On how *the motion of the stars* changed the language of science: a corpus-based study of deverbal nominalizations in astronomy texts from 1700 to 1900

Iria M. Bello Viruega

Doctoral thesis

2014

Supervised by
Dr. Begoña Crespo García

Departament of English Philology

University of A Coruña (Spain)

UNIVERSIDADE DA CORUÑA

Acknowledgments

I would like to express my gratitude to my supervisor, Prof. Begoña Crespo. From my time as an undergraduate student, curious about the history of English, she has walked next to me and has carefully guided my progress. Her patience confronting my growing doubts and anxieties has reassured my academic development and her domain expertise has given perspective to the whole project. I would also like to thank all the members of Muste. This doctoral thesis would have never been carried out without the incessant determination of Prof. Isabel Moskowich to build the corpus. Similarly, I am very grateful to the rest of the compilers, Prof. Inés Lareo and Prof. Gonzalo Camiña, because of all the impressive time and effort they invested building the corpus, as well as for serving as role models for my own academic development. I also thank the rest of the members of the research group, especially Paula, María and Luis, for sharing with me all those days in our beloved musty basement, and Fanny, for helping me endure with a laugh the downsides of research in Cambridge.

I am also very grateful to Prof. Laura Wright, who welcomed me at the English Department in Cambridge University. Working in such an important place for the history of English and having direct access to the first editions of some of the most influential books in the development of the scientific register meant the fulfillment of a dream. At the University of A Coruña, I would like to thank all the professors that have helped me develop my interest in linguistics. I want to highlight Prof. Manuel Fernández for his kind words and for introducing me in the field of cognitive linguistics and Prof. Margarita Alonso and the LYS research group for giving me the chance of getting acquainted with the study of collocations.

My gratitude is also extended to all my friends. I am very grateful to all my linguist friends for all our valuable debates, which have shaped my understanding of language: to Deeni and her insights on Mexican Spanish, which made realize about the importance of standard varieties; to Carmen and her master classes on Chinese linguistics to Orsi and our debates on collocations (and Chinese linguistics); to Max and his valuable insights on coaching and foreign language teaching. I would also like to thank Lola for being my editor and for her enjoyable disrespect for linguistics, as well as Rosa for her attempts to teach me statistics, and Elena for her support in our early years as graduate

students. To all my other friends outside the academia, I would like to thank all their support and their attempts to show interest whenever I started to digress about my research.

Finally, I would also like to thank my family for their love and support in all my years of education. I am very grateful to my parents for granting me complete freedom in choosing my path. I want to thank my mother for serving as a source of life inspiration and especially for instilling in me a passion for knowledge and self-amelioration. I would also like to thank my sister Amil because of her constant support as well as for believing in me especially when I did not do it.

To my parents and my sister.

Resumen

Esta tesis doctoral supone un análisis sobre las nominalizaciones deverbales formadas por sufijación en textos de astronomía escritos en inglés en los siglos XVIII y XIX. El material de análisis para este estudio fue tomado del *Corpus of English Texts on Astronomy (CETA)* (Moskowich et al., 2012). El corpus contiene dos textos por década y cada una de las muestras contiene alrededor de 10.000 palabras, lo que hace un total de 400.000 palabras analizables. El objetivo principal de este trabajo es el estudio de las nominalizaciones como marcadores del discurso científico en inglés moderno tardío. Varios cambios sociales que tuvieron lugar a principios de la Europa moderna afectaron gravemente el enfoque científico y esto tuvo un efecto directo en su lenguaje. Para llevar a cabo el análisis, he creado una tipología de las nominalizaciones que tiene en cuenta características formales y funcionales. Se formularon una serie de variables independientes: por un lado, las variables extralingüísticas abarcaron la cronología, el sexo del autor, el lugar de educación del autor y el tipo de texto; el resto de variables lingüísticas abordaron la estructura de las nominalizaciones y sus frases nominales e incluyeron el uso de sufijos, la etimología, los modificadores, las construcciones posesivas, la inclusión de agentes y circunstancias y la función sintáctica. Estas variables se aplicaron primero al número total de nominalizaciones encontradas en el corpus (8.446) y luego a las cuatro tipologías creadas.

Resumo

Esta tese de doutoramento é unha análise sobre as nominalizacións deverbais formadas por sufixación en textos de astronomía escritos en inglés nos séculos XVIII e XIX. O material corpus para este estudo foi tirado do *Corpus of English Texts on Astronomy (CETA)* (Moskowich et al., 2012). O corpus contén dous textos por década e cada mostra contén aproximadamente 10.000 palabras, o que fai un total de 400.000 palabras analizábles. O principal obxectivo deste traballo é estudar as nominalizacións como marcadores do discurso científico en inglés moderno tardío. Varios cambios sociais que se produciron en Europa ao comezo da etapa moderna afectaron severamente os enfoques de cara a ciencia e iso tivo un efecto directo sobre a súa linguaxe. Para realizar a análise creei unha tipoloxía de nominalizacións tendo en conta características formais e funcionais. Tamén formulei unha serie de variables independientes: por unha banda, as variables extralinguísticas inclúen cronoloxía, sexo do autor, lugar de educación do autor e tipo de texto; por outra, unha serie de variables intralingüísticas cubren aspectos relacionados coa estrutura das nominalizacións e as súas frases nominais. Estas incluen o uso de sufixos, a etimoloxía, os modificadores empregados, as construcións posessivas, a inclusión de axentes e circunstancias e a función sintáctica. Estas variables foron primeiramente aplicadas ao número total de nominalizacións atopadas no corpus (8.446) e, a continuación, as catro tipoloxías creadas para este estudo.

Abstract

This doctoral thesis analyzes deverbal nominalizations formed through suffixation in astronomy texts written in English in the eighteenth and nineteenth centuries. The corpus material for this study was taken from the *The Corpus of English Texts on Astronomy (CETA)* (Moskowich et al., 2012). The corpus contains two texts per decade and each sample text contains around 10,000 words, which makes a total of 400,000 analyzable words. The main aim of this work is to study nominalizations as scientific discourse markers in late Modern English. Several social changes that took place in early Modern Europe affected severely approaches to science and this had a direct effect on its language. To carry out the analysis a typology of nominalizations acknowledging formal and functional features was created and independent variables were formulated: on the one hand, extralinguistic variables included chronology, sex of author, place of education and text-type; on the other hand, intralinguistic variables dealt with the structure of nominalizations and their NPs and included suffix use, etymology, modifiers, possessive constructions, agency and circumstance inclusion and syntactical function. These variables were first applied to the total number of nominalizations found in the corpus (8,446) and then to the four typologies created.

Table of contents

List of tables

List of figures

Abbreviations

AP	Adjective Phrase
CC	*Coruña Corpus*
CCT	*Coruña Corpus Tool*
CECheT	*Corpus of English Texts on Chemistry*
CHET	*Corpus of History English Texts*
CELiST	*Corpus of English Life Sciences Texts*
CEMaT	*Corpus of English Mathematics Texts*
CETA	*Corpus of English Texts on Astronomy*
CEPHiT	*Corpus of English Philosophy Texts*
CETEL	*Corpus of English Texts on Linguistics*
CETEPH	*Corpus of English Texts on Physics*
N	Noun
N.A.	Non Applicable
NP	Noun Phrase
ICM	Idealized Cognitive Model
OE	Old English
OED	*Oxford English Dictionary*
PP	Prepositional Phrase
SFL	Systemic Functional Linguistics
SVO	Subject Verb Object
TGG	Transformational-Generative Grammar
U.K.	United Kingdom
U.S.	United States
V	Verb
VG	Verb Group
VP	Verb Phrase

Introduction

Change your language and you may end up changing your thoughts. Change your thoughts and you may end up changing your language. Science and the language of science are two indissoluble entities. The difficulty of scientific language is not limited to the lexical level, as it also applies to a range of specific grammar structures (Halliday, 2004). Nominalizations, one of the most well-known scientific discourse markers, are complex structures that encode processes into nouns. When nominalizing, semantic information is usually left out, which increases the degree of ambiguity and the difficulty of correctly decoding the sentence. Studying the evolution of this process across time may help understand its present-day situation. Halliday (2004, p. 105) asserted that the evolution of the language of science in the last 400 or 500 years has developed new ways of nominalizing. The object of research of this study lies more specifically on deverbal nominalizations formed by suffixation in scientific register texts in English written in the eighteenth and nineteenth centuries. In this time-span, several social changes affected approaches to science and this had its effects on language. Empiricist scholars adopted stylistic guidelines regarding morphosyntax,

specialized vocabulary and text structure for the presentation of the papers of observational or experimental nature which became a standard of writing after the founding of the Royal Society (Crespo, 2004a, 2011). Nominalizations became one of these distinctive markers and consequently its complexity and frequency have progressively increased ever since.

Natural languages have multiple resources to express similar ideas. This flexibility, which may entail shades of meaning, has attracted the attention of linguists and speakers, who see there the richness of language. Nominalization is understood as a linguistic expression of a conceptual representation of a process or state of affairs in a nominal form. According to Downing (1997, p. 147) situations and processes can be expressed through nominalizations, as in (1)

> (1) From whence it is gathered, that **the apparent progreſſive Motion of the Fixed Stars** hath gone forward one Degree towards the conſequent Signs, in about Seventy Years ſpace (Whiston, 1715, p. 14; emphasis added)

or through finite sentences, as in (2)

> (2) Aſtronomers know that not only the 12 Conſtellations of the Zodiac, but alſo **all the fix'd Stars move from the Weſt toward the Eaſt** about 50" in a Year, or one Degree in 71 Years, in Circles parallel to the Ecliptick (Watts, 1726, p. 34; emphasis added)

Obviously, although in general terms they convey the same meaning, each of these two linguistic encodings have a different structure and fulfill different functions in texts. In (2) *move* controls the syntax of the whole sentence through a system of obligatory valencies and optional adjuncts. Thus the agent (*Conſtellations of the Zodiac, but alſo all the fix'd Stars*) is expressed in the subject. The direction of the movement

(*from the Weſt toward the Eaſt*) is made explicit and we are also given information on how (*in Circles parallel to the Ecliptick*) and how much (*about 50" in a Year, or one Degree in 71 Years*) constellations and stars move. Similarly, in (1) *motion* also exerts control over its phrase but it is inserted into a larger sentence. Structure is not so rigid in this case as, by definition, all elements in the NP with exception of the Head are optional. This allows a more complex arrangement. Thus, in (1) information about the agent (*of the Fixed Stars*), and how (*progressive*) the motion is performed is contained in the post- and premodifying fields of motion, respectively. In this particular example, additional information about the process is also found in the VP: as in (2), we are also given information about how much the stars moved (*one Degree*) and in what direction (*towards the conſequent Signs, in about Seventy Years ſpace*).

Functionally, the expression of the process as a verb in (2) may be nearer the speaker's experience in terms of chronological sequencing and experience of reality. All the information about the process is kept near the verb. However, the configuration of (1) allows the presentation of the process of *moving* not as a simple account of reality but as a reified consequence of previous discourse. To illustrate this point, it may be necessary to have a look at the context in which (1) was inserted:

> (1a) Beſides this general apparent **Motion**, which is perform'd in well nigh the Space of 24 Hours, in a perpetual Succeſſion, the Fixed Stars ſeem to be **moved** with another **Motion** alſo almoſt quite contrary to the former; for they are found to change and enlarge their Longitudes, that is, their Diſtances from the beginning of Aries, reckoned along the Ecliptic, or towards the conſequent Signs. For what Fixed Stars appeared in Hipparchus's, or even Ptolemy's time, in that Dodecatomorium, or Twelfth Part of the Zodiack called Aries, appear now in the Sign Taurus. What Stars were reckoned in time paſt as belonging to Taurus are now aſcribed to Gemini, and ſo on: From whence it is gathered, that **the**

> **apparent progreſſive Motion of the Fixed Stars** hath gone forward one Degree towards the conſequent Signs, in about Seventy Years ſpace; and that with an even Velocity (Whiston, 1715, p. 14; emphasis added).

The position of *the apparent progreſſive Motion of the Fixed Stars* at the end of the paragraph, together with the first part of the sentence in which it is inserted (*From whence it is gathered, that*) may point out that the process presented in motion is not a changeable situation but rather a reified element that serves as a recapitulation of the contents previously presented. Not all nominalizations introduce this type of relationship. However, they generally serve as functional guidelines which help organize information in the mind of the reader.

The study of nominalizations has attracted the attention of numerous scholars. In English interest started within the generativist school back in the 1960s with Lees' (1960), Chomsky's (1970) and Newmeyer's (1971) works. With these works nominalizations gained momentum and became very important within this school. In this line, we can find the work of Grimshaw (1990), Jackendoff (1975), Hazout (1995), Siloni (1997) and Zucchi (1993). In the 1980s functionalism tried to describe the functions and advantages involved in the use of nominalizations. Thus Albentosa (1997), Banks (2003, 2005a, 2005b), Guillén (1998), Halliday (1985), Ravelli (1988) and Ventola (1996) provided their theories and studies about what they called grammatical metaphor. In recent years interest has arisen in the field of computational linguistics for the treatment and semantic representation of nominalizations. Thus, the studies provided by Eberle Faasz and Ulrich (2011), Girju et al. (2009), Hull and Gómez (2000), Lapata (2002) and Padó, Pennacchiotti and Sporleder (2008) to name just a few can account for this renewed interest in nominalizations.

Some authors, like Billig (2008), van Dijk (2006), Fairclough (1992), Fowler (1991) and Wodak and Meyer (2001) have focused on the relationship between nominalizations, ideology and power and have based their studies on journalistic texts. Others have directed their attention to the scientific register and have analyzed the origins, evolution and use of nominalizations as discourse markers. In this vein of study we can find the works of Albentosa and Moya (2000), Banks (2001, 2003, 2005a, 2005b, 2007, 2008), Halliday (1985), Halliday and Martin (1993), Sušinskienė (2004, 2009, 2010a, 2010b, 2012) and Vázquez (2006). This doctoral thesis joins this track of research: the study of nominalizations in scientific register written in English in the late Modern period. The choice of astronomy was made based on the premise that it was perhaps one of the scientific disciplines that best epitomized the change of direction Western science took after the Scientific Revolution. Acknowledging a long, solid history as scientific discipline dating back to Antiquity, astronomers in the seventeenth century had the opportunity of using the latest technical advancements of their time to improve their understanding of the sky. Recommendations on epistemological methods emphasizing observation and experimentation and the upgrading of Newton's teachings supposed a real revolution within the discipline. Since astronomers in the late Modern period changed their thoughts about science, I expect to find changes in the language as well.

The aim of this doctoral thesis is to carry out an analysis of deverbal nominalizations formed by affixation in a corpus of astronomy texts written in English in the eighteenth and nineteenth centuries in order to determine to what extent these linguistic features organize discourse and can be considered discourse markers. This line of research is part of the study of the historical development of English for specific

purposes (ESP). This main objective is broken down into five objectives outlined below:

1. **Study of nominalizations as scientific discourse markers**: nominalizations are complex structures that present processes in a nominal form. Agency and other information about the process is usually excluded and that enhances the degree of ambiguity in texts. This study will analyze in depth not only the morphosyntax but also the functional implications of nominalizations in texts, especially with regard to their role as scientific discourse markers.
2. **Corpus analysis of nominalizations according to diachronic criteria**: following this guideline I will trace the evolution of nominalizations during the two-hundred years covered in the study.
3. **Corpus analysis of nominalizations according to extralinguistic variables**: these include chronology, sex of author, place of education and text-type. These variables can help place the relationship between language and society in its historical dimension. Similarly, its application to data results will provide information about how sociological factors can cause linguistic variation.
4. **Corpus analysis of nominalizations according to linguistic variables**: the study of the structure of nominalizations is carried out at two levels: an etymological analysis of both roots and suffixes will cover the lexical level. Since nominalizations always function as heads of an NP, all the elements of the phrase are also analyzed with the underlying expectation that they will include information about the process. Finally, syntactic analysis will shed some light on functions within texts.
5. **Corpus analysis of nominalizations according to typology**: this will help clarify that nominalizations are part of a continuum. I believe that processes may acquire a

certain degree of semantic and lexical features of either nouns or verbs when expressed linguistically as finite sentences or nominalizations. A typology was developed following this claim and one of the objectives of the study was the analysis of the different typologies after the application of several linguistic variables.

6. **Socio-historical analysis of science in the modern period**: even if this is not one of the main objectives and it was originally conceived as a way of complementing corpus analysis, the study of science in the modern period and especially the situation of exclusion of women scientists became one important focus of analysis.

In figure 1 the methodology used for obtaining data from the corpus, carrying out the analysis and obtaining conclusions, is presented. Obviously, and although it is not very well represented in the figure, a great deal of revision and reconsideration of bibliographical sources was needed in all stages.

Before searching the corpus, it was necessary to establish the definition of nominalization to set the boundaries. After a detailed study of the bibliography on nominalizations, I limited the scope of this study to lexical nominalizations formed by suffixation expressing a process. As a result, the establishment of a set of criteria for disambiguating those nominalizations included for analysis had to be envisaged[1]. Since it was agreed that only deverbal nominalizations formed by suffixation were considered for analysis, a list of deverbal suffixes was also set. This list was then used to perform the queries in the *CCT*. The wildcard (.*)[2] was combined with the strings of letters of

[1] These disambiguation criteria are fully explained in section 3.2.

[2] "An asterisk represents (*) "the preceding character, zero or more times". Used after a dot (result.*) it expands the search to encompass any word containing the preceding form (result), followed by something else: resulting, results, resulted, etc." (*Coruña Corpus Manual*).

the chosen suffixes. The addition of <s> to all strings of letters covered up for plural forms and resulted in fourteen spreadsheets.

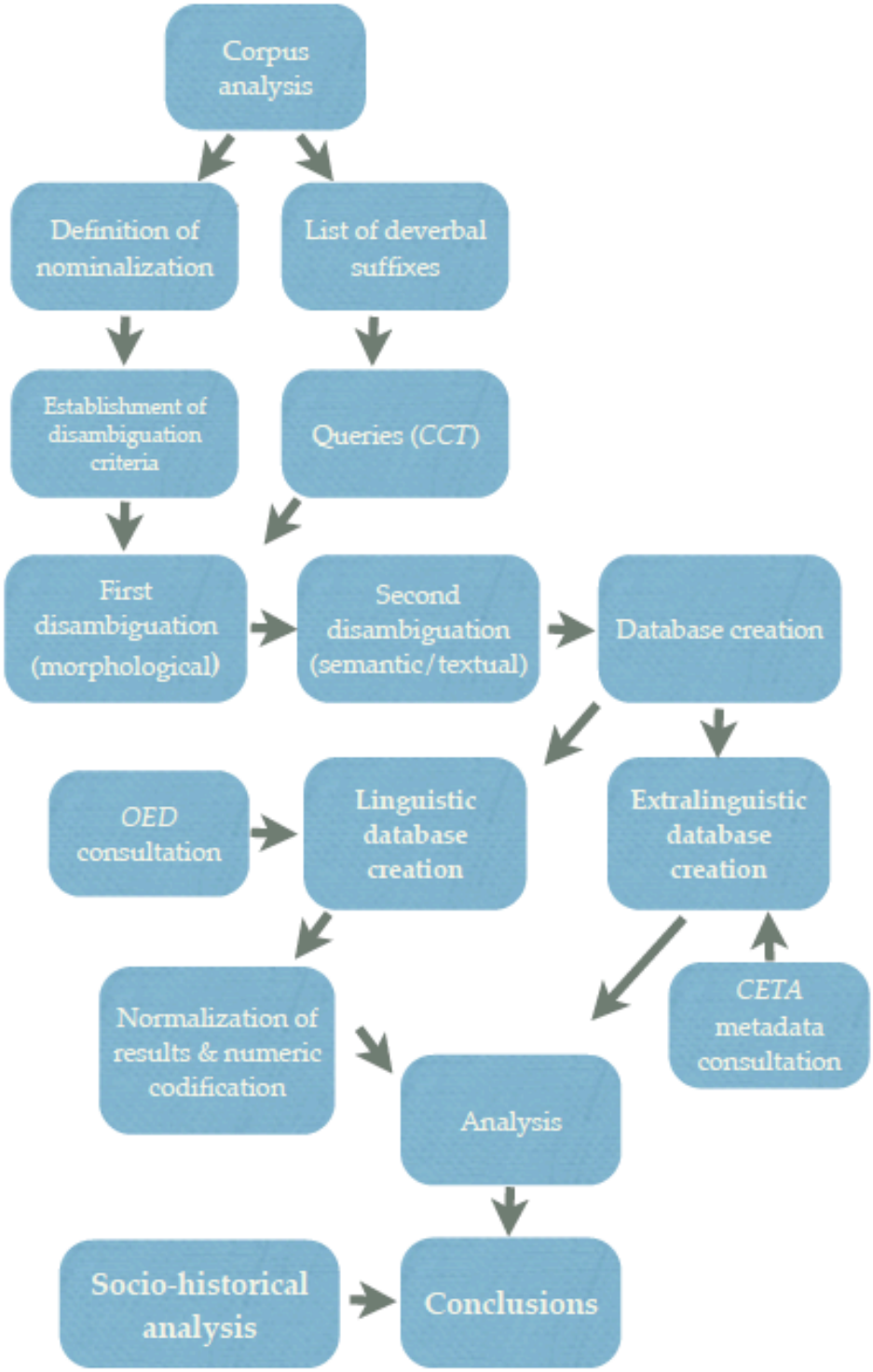


Figure 1: Methodology.

Results were then subjected to manual disambiguation, as the *CCT* does not recognize suffixes but strings of words and the concordances generated included also gerunds and other words ending with these letters. The first disambiguation was based on word class criteria and eliminated those words that were not nouns. After this, some occurrences were marked as ambiguous. Context reading and semantic disambiguation were carried out at the second stage to sort out the final number of nominalizations considered for study (8,446). After that, all effort was drawn to database creation. The database containing the extralinguistic variables was filled out with the data found in the metadata files in *CETA*. The database with the linguistic variables required more time and effort. Also, *OED* was consulted for etymological variables. Once results were normalized and numerically codified, the analysis was carried out according to the variables presented in section 3.2. The conclusions of this study will include not only the results of corpus analysis but also a reading of the socio-historical analysis carried out in chapter one.

This thesis is divided into four main chapters, which are complemented with an introduction and a chapter of conclusions. These four chapters deal in depth with both extralinguistic and linguistic features that had taken part in linguistic change and shaped the use of nominalizations in scientific texts written in English between 1700 and 1900, that is, the main object of study.

Chapter one aims at introducing the context of production of the texts. Science and the language of science are two indissoluble entities (Halliday, 1985) therefore, to understand one, we must study the other as well. This is the general motivation for this chapter, in which a brief account of science and the scientific register in the modern period is provided. Attention is drawn to three areas: a) the general situation of science

in the period, which includes an analysis of the Scientific Revolution that took place in the seventeenth century and how it affected approaches to science and the places where it was performed. An account of the situation of women scientists in the modern period, encompassing and analysis of their situation in the academia, their consideration for intellectual activities and their education is also provided. The fact that the number of texts written by women in the corpus is reduced –only 5% of the total– is a reflection of the situation of exclusion in the period; b) a more in-depth analysis of the situation of astronomy in English-speaking countries in the eighteenth and nineteenth centuries, which will enable a deeper understanding of the extralinguistic variables of study; c) a brief analysis of English in the modern period, which includes a special mention to the creation of the scientific register in English.

Chapter two contains a review of the literature on nominalizations as well as a description of their main theoretical considerations. The main theories about nominalization provided by different linguistic schools are discussed. This includes an analysis of the concept of transference, which is considered key in many theories. Morphosyntax is approached from two levels: formation and semantics belong to the lexical level, whereas at the phrasal level, nominalizations always function as heads of a NP. A great deal of information about the process expressed in the noun is contained in the elements of the nominalization NP, so an analysis of their structure, semantics and function is extremely pertinent. Additionally this chapter contains an interpretation of the main functional implications usually associated with nominalizations and a reading of them as scientific discourse markers. The final pages of the chapter are devoted to the description of the four typologies that were created for this study.

Chapter three describes in detail the corpus of texts used for the analysis as well as the methodology used in it. In the first part the corpus used for analysis, the *Corpus of English Texts on Astronomy* (*CETA* henceforth), and the search engine, the Coruña Corpus Tool, (*CCT* henceforth) are presented. The description of the corpus, the lengthiest part of the chapter, contains an analysis of its general features and its parts. Issues like the treatment of texts in the corpus, the metadata files and some information about prologues are included. Concerning methodology, the disambiguation criteria and the variables of study are here presented.

Chapter four constitutes the analysis of the data obtained after corpus exploitation. After the application of the variables described in chapter three, the results are discussed here. The analysis is attempted from two different angles. First all nominalizations are confronted with extralinguistic and linguistic variables. Then, a linguistic analysis of nominalizations according to the four typologies described in chapter two is carried out. The application of extralinguistic variables to typologies did not show significative results so I have preferred to exclude it from this chapter so as not complicate the presentation of results. The application of linguistic variables, both to the whole number of nominalizations and the typological analysis follows again the lexical/phrasal distinction. Thus, the analysis of the nominalization as a lexical unit is concerned first with its morphology and then with its function. Then, the analysis of nominalization NPs also emulates the form/function distinction, as it includes not only a study of the elements but also of the functions and semantics fulfilled by those elements within the NP governed by the nominalization.

Finally, in chapter five the overall conclusions and the future lines of work are offered. My aim is that the conclusions here presented will summarize the data

contained in chapter four by meeting the requirements laid out in previous chapters. The ultimate goal is that the basis provided by the quantitative data obtained from corpus analysis will help account for the development of nominalizations as scientific discourse markers in English in the eighteenth and nineteenth centuries.

1. Science and language in the modern period

It is impossible to separate the history of a language from the external history of the speakers of that language (Labov, 1966; Millward, 1988). As Halliday (1985) pointed out, science and the language of science are two indissoluble entities. The language of science and general language cannot be broken apart either, it could be argued. Any major change in the lives of people will have its reflection on language. It is therefore important to pay attention to extralinguistic features, especially in the case of a diachronic study in which our assumptions about science or scientific writing may not coincide with the reality of English two hundred years ago[3]. All the information provided in this first chapter will then be used to interpret the data resulting from the linguistic analysis. The ultimate goal is to be able to explain how and in what way nominalizations developed using linguistic and extralinguistic information.

[3] As Camiña (2013) pointed out, in a linguistic study about the scientific register of past historical periods, whiggism, that is judging the past in terms of the present (Henry, 2002, p. 2) could be an easy mistake. Thus, even if the word 'science' was coined in the fifteenth century (*OED online*), until the nineteenth century, scientists were called 'natural philosophers' and they attempted to explain physical phenomena in a variety of interdisciplinary ways. The changes of the Scientific Revolution were so profound that implied the coinage of the word 'scientist', which did not exist before. It was William Whewell, who, under the petition of the poet Coleridge, coined the word in 1833 as a replacement for 'natural philosopher' (Camiña, 2013, p. 30). Despite these differences, the terms 'scientist' and 'natural philosopher' have been treated as synonyms in this study in an attempt not to distract attention from the main topic of study: the language used by these people.

This chapter is divided in four sections. It starts with an account of the state of science in the eighteenth and nineteenth centuries (section 1.1). Even if the Scientific Revolution is commonly accepted to have taken place in the seventeenth century –that is, one century before the starting point of the corpus' time span– a brief analysis of the scientific method is provided here (section 1.1.1) because the scientific practices and the scientific language used in the eighteenth century derive directly from the theoretical framework established in the previous century. The institutionalization of science and the academy movement are also analyzed (section 1.1.2) because they are considered crucial factors for the establishment of a scientific register. The final pages of the section are concerned with the role of women scientists in these centuries (section 1.1.3).

After these three initial subsections, a brief account of the situation of the discipline of astronomy will be provided in section 1.2. I consider it important to possess a vague understanding of what the content of the texts may be and thus, this section is aimed at outlining the research interests and the evolution of the discipline until the modern period. The last section of this chapter (1.3) deals with English in the modern period as well as with the springing and evolution of English as a language of science (section 1.3.1). Again, the only aim of this section is to provide a framework for the analysis of nominalizations used in astronomy texts in late Modern English.

1.1. Science in the modern period

The modern period was crucial in the development of science. Since Ancient times, there had been no reformulations. All natural philosophers were concerned with the application of old patterns in their study. It is not until the seventeenth century that the method itself was put into debate. Several historical events contributed to this change. The fall of Constantinople in 1453, usually cited as one of the stepping stones that led to the modern period, eventually transformed knowledge in the western world and helped shape the understanding of modern science (Camiña, 2013, p. 28). Columbus' discovery of the New World in 1492 also triggered a navigation fever and leading European countries were eager to sponsor those who researched and developed instruments to improve navigation. As in other periods of history, all these changes were a cause and a consequence of a series of technological advances. In the case of modern science, the use of the compass[4] and the invention of the printing press (1450) and the telescope (1602) can be considered crucial. These political and technological changes was also paired with the splendor of the Renaissance, a movement that turned the political turmoil of the age into a reconsideration of cultural practices. The humanist movement that drew attention on the individual served then to shift scientific methodology from contemplation to experimentation. This section deals with this crucial change under three different prisms: section 1.1.1 is devoted to the origins of the scientific method itself, section 1.1.2 deals with the academy movement, one of the most visible consequences of the reformulation of science. Finally, section 1.1.3 is concerned with the role of women scientists in the period.

[4] The compass had been invented in China centuries ago but it did not became popular in Europe until the fifteenth century.

1.1.1. The (r)evolution of science and the scientific method

The Scientific Revolution that took place in Western science in the seventeenth century[5] meant the foundations of some decisive conceptual, methodological and institutional changes that affected the development of Western science history (Henry, 2002, p. 2). The term "revolution", however, should be carefully assessed as, according to some scholars (Crombie, 1974; Hall, 1954), this movement did not aim at breaking up with the past and was rather a natural evolution that started in the Middle Ages with philosophers like William of Ockham, who maintained that real knowledge came from sensation (Crombie, 1974, p. 217; Hall, 1954, p. 163). Seventeenth-century empiricism shared this basic principle but with a slightly different approach, especially regarding the application of science to the solving of problems. Basically, the new philosophy was a reaction against old Aristotelian methods of learning. Embedded in its own principles, the Aristotelian tradition failed in opening up new directions and was stuck in a series of abstractions and repetitions. As Hall (1954: 165) explained "Aristotelian science was a hollow structure, dealing with abstractions rather than real things, justified by no fertility in works."

From medieval times, the Catholic Church had the monopoly in the keeping and transmission of knowledge. By the beginning of the nineteenth century science was available to most social classes and scientists and researchers were not unusual jobs.

[5] The peak of the movement took place in mid-seventeenth century. However, precursor activities already started in the sixteenth century and the eighteenth century meant a consolidation of the movement (Camiña, 2013, p. 29).

Nevertheless, this change was slow and many factors took part in it. Political issues modified the distribution of power and the people that had the control of knowledge. Social changes led to different approaches to knowledge and the creation of different schools of thought, which ultimately led to a shift from insight to systematic study (Burke, 2004). In the case of England, the humanist movement that had reached the country in the last decades of the fifteenth century meant a development of experimentation as a method of obtaining knowledge. As Crespo (2004a) pointed out, previous scholastic models of knowledge were based on the establishment of a series of reasoned, purely theoretical deductions derived from a set of previously established principles. Humanists, on the other hand, were rather concerned with searching solutions to specific problems. Their interest was not on immutable divine truths but on specific issues that could affect man. Another implication of the humanist movement was the universalization of knowledge, which meant that knowledge became accessible outside universities and religious centers.

The names of Francis Bacon (1561-1642), Galileo Galilei (1564-1642) and René Descartes (1596-1650) have been traditionally cited as the leading precursors of the new philosophy of science (Goodman & Russell, 1991; Hall, 1954; Krämer-Friedrich, 1988) The British Francis Bacon thought that knowledge should be reformed. Continuous circular communication was causing an endless circle of error repetition in the sciences. To correct this situation, he suggested that books should be left outside and new information should be sought through the direct observation of nature. Observation, collection and classification of natural phenomena became thus the main endeavors of natural philosophers. The innovative approach to science introduced by Bacon has been repeatedly praised (Camiña, 2013; Snyder, 2009) and it has been claimed that “modern

science was consciously modeled upon Bacon's system" (Hall, 1954, p. 166). Bacon's contribution to the rise of science in the seventeenth century can be summarized in three basic points: a) his writings and vehement style served as an impulse for science; b) his development of the theoretical basis of the inductive method, based on observation and experiment to extract conclusions became the standard methodology; c) his emphasis on the separation between science and religion became a reality. This proposal, however, was far from being rebellious. His idea was that by knowing more, people would be able to appreciate better God's work and His Glory would be revealed. Baconianism was indeed very well regarded by followers of Oliver Cromwell during the Interregnum (1642-60). Coley (1991) asserted that the rise of Puritanism was closely linked to the rise of this new approach to learning. Self-restraint, orderliness and simplicity, the main tenets of Puritanism, greatly coincided with the new science's necessities.

Another precursor of the new philosophy of science, Galileo, was also concerned with specific problems. For him, the most important step was abstraction: real properties of bodies are purely physical and there is no distinction between real and mathematical truth. He was also rather concerned with the classification of science and tended to assume that labeling can be misleading if based on superficial characteristics. However, not all scientists shared his view. For Descartes all knowledge had to come through deduction. Experimentation was thus an invalid way to reach valid conclusions. As Hall (1954, p. 173) remarked "the difference between Galileo and Bacon in this respect is that the former emphasized mainly the role of experiment in testing a theory, or determining its constants, while the latter stressed the role of experiment as a means of obtaining information." As it is well-known, even if their theories were extremely influential, neither Galileo nor Descartes became the only sources of knowledge. The

idea that the Scientific Revolution and the experimental method was a solid theory is partly wrong. In the beginning there were many discrepancies and on the whole, the idea of a unified Scientific Revolution was constructed time after it actually took place.

As far as scientific schools are concerned, the seventeenth and eighteenth centuries witnessed a constant dispute between empiricism and rationalism. Empiricists and rationalists did not agree on the validity of purely sensorial or deductive methods when applied to propositions related to the external world. The theory of empiricism, first formulated in the seventeenth century highlighted the role of experience and sensation as opposed to the traditional process of reasoning inherited from medieval scholasticism. Empiricists were concerned with the establishment of a method that could provide specific solutions for particular problems. This practical view of knowledge rejected the validity of intuition and deduction as valid ways of knowledge. Empiricist philosophers –Locke, Hume, Berkeley– found it necessary to directly expose our senses to the subject of study. Empirical study, thus, became a synonym for a kind of study that depends on evidence that may be noticed by the senses. Experiments were required hence to formulate valid scientific statements.

Following the teachings of Bacon, Galileo and Descartes, Isaac Newton (1643-1727) is one of the leading figures of the Scientific Revolution. He is considered to have revolutionized modern science and the practical application of his theories concerned scientists well until the twentieth century. The success of Newtonianism can be said to rest upon several issues. First, the method itself presented an innovative approach to science in a time where Aristotelian scholasticism was indeed stuck and could not keep up with changes in society. Nevertheless, no matter how innovative Newtonianism was, “neither Newton nor the seventeenth century at large invented the

concept of experimentation", Applebaum (2000, p. 461) claimed. The recipe for success should, hence, come from other sides. The propaganda carried out by members of the scientific academies, the main materialization of the Scientific Revolution, was very important in this respect, as will be shown in the following section. New philosophers also secured an affinity with religion. This was especially important, given that religion had been the main bastion of science for more than one thousand years. An abrupt confrontation with religious authorities would have minimized the expectations of success for the new science. The Baconian scientific method was, as has been already pointed out, deeply religious, even if it sought for a separation of science and religion. Similarly, Gresham College, one of the main precursors of the London Royal Society, was a Puritan institution. The allegiance to the Crown was another decisive point in the success of Newtonianism. Obviously looking for funding, new philosophers made sure that the network of academic societies and the work carried out there was made available to the state. Finally, the utilitarian approach that was included in the new philosophy helped establish economic ties in society. Commercial applications of scientific discoveries together with the growing market of increasingly specialized scientific apparatus could be pointed out as indicators that this revolution was not only scientific but also economic.

This new idea of modern science, though highly innovative, still relied too much in Aristotle and other traditions. Magic, astrology, alchemy and witchcraft were still very important. The culture of curiosities displayed in public, performances and experiments were also part of the new science, as Bensaude-Vincent and Blondel (2008) claimed. Our present-day understanding of sciences is narrower than it was for modern Europe population. Popular courses in chemistry, botany and physics were common. A

curious audience with amateur knowledge in those disciplines demanded experiments to satisfy their curiosity, which sometimes became entertainment shows. Instrument makers also favored popular representations of science in an attempt to raise sales. For Bensaude-Vincent and Blondel (2008), entertainment is not incompatible with pedagogy and these shows played a double role: they were fun demonstrations serving as disseminators of science. Public demonstrations attired the attention of a wider audience. The benefits of this were numerous, similar to the workshops and activities organized by today's museums and scientific institutions: they were fantastic disseminators of science, created a public taste and contributed to the development of science and instrument-making. Ultimately, they also contributed to the advancement of society.

As far as the eighteenth century is concerned, the Enlightenment is considered the most important historical, intellectual movement in Europe and America at that time. The Age of reason aimed at establishing an authoritative system which would manage to organize society and banish superstition and irrationality. Hall (1954, p. 216) depicted this tension between English scientific groups, followers of Newton and French and German Cartesians, led by Leibniz. He pointed out that this antagonism was especially evident between 1665 and 1720 and it triggered allegiance based on nationalism rather than scientific reasons. At the same time, he also clarified that the fundamentals of the scientific community were not at risk because, after all, both schools of thought were in the same direction and they were not completely opposites.

The increase of industrial production contributed to the development of a type of science which searched to be applicable to situations of real life. Deduction and intuition were thus substituted by practical applications of scientific theories. Although

the formal distinction between pure and applied sciences was not established until some time later, the eighteenth century was the turning point which marked the beginning of the applied branches of science. Applied sciences were seen as the best way to improve the development of industry. Consequently, science acquired some social implications and it was generally considered as a form of culture which would help improve the national manufacturing, agriculture, medicine, administration and other fields of society.

The nineteenth century implied, on the other hand, the professionalization of science. This century was characterized by a constant search for progress. The effects of the Industrial Revolution had already established by that time the new courses that science had to follow. This was an age of great development for the applied branches of science. Basic elements of our everyday life, such as the automobile, the airplane or the telephone were invented in this century. It is also important to highlight the key role that Darwin's theories played in nineteenth-century British society. Apart from raising social controversy, they changed the general perspective of science in modern society, which became increasingly more differentiated from religion. Although Darwin was mainly a biologist, his theories reformulated the whole view of science within society. Almost two centuries after the beginning of the Scientific Revolution, the panorama of science in the Western world had radically changed. From complicated manuscripts based on Aristotelian theories and written in isolated monasteries, science had become a profession with a rigid method based on observation and experimentation and whose influence had a direct impact on society at the end of the nineteenth century. This change would have not been the same if it had not been paired with an institutionalization process that physically secured the visibility of the new people of science.

1.1.2. The Scientific Revolution and the institutionalization of science.

Scholars have remarked that one of the main consequences of the Scientific Revolution was the establishment of scientific academies as the leading organizers and disseminators of modern science (Applebaum, 2000; Bartholomew, 1991; Brush, 1988; Burke, 2000; Coley, 1991; Feingold, 1989; Hall, 1954; Roberts, 1991; Russell, 1991; Voss, 1980). Originated as an alternative to universities, which were tied to old scholastic approaches to science, academies are a modern product. In spite of its present-day reputation, defenders of this organization of science had to fight for a market niche. The networking system of scientific academies and institutions originated in the second half of the seventeenth century and in less than one century it had virtually taken over the practice of science. After all, academies were –and still are– a state-favored mechanism to control society. In his article about scientific academies in Europe in the eighteenth century, Voss (1980, p. 44) commented

> Die Akademie ist ein Produkt der modernen Wissenschaftentwicklung. Sie ist gerade aus dem Widerspruch gegen die ältere, sogenannte mittelalterliche Wissenchaftsüberlieferung entstanden. Der Protektor der Akademien, also Staat oder Krone, musste ihnen einen Spielraum und eine Bewegungsfreiheit gewähren, die weder bei den damaligen Universitäten bestand noch in der Praxis des absolutistischen Staates irgendein Vorbild hatte.

The academy movement soon spread all over Europe and all scientific disciplines. Traditionally, astronomy, physics and mathematics are considered to be the pioneering disciplines but chemistry, botanic, mineralogy and natural history had their momentum at the end of the eighteenth century and reached an age of accolade in the course of the

following century. Humanist disciplines were slightly left behind. Philology was perhaps the luckiest discipline and studies in classical philology proliferated in the nineteenth century[6].

Italy was the initiator of the academy movement in the early years of the seventeenth century. As is well-known, this country had been a pioneer in the exportation of the Renaissance to all Europe. The first modern scientific academy was established in 1609 by Prince Federigo Cesi (1585-1630). The Accademia dei Lincei, which lasted for more than twenty years, had Galileo as one of its most eminent members. Cesi, the alma mater of the project, made sure that philosophers in his accademia shared with him a new attitude to science, based on an explicit emphasis on observation and experiment and desertion of old scholastic models. The name of the academy comes from lynx and it reflects their attitude about learning: the members of the academy knew that what we cannot see can indeed exist, as the lynxes and eagles, which are known to see more than the human eye could distinguish. The lynceans were mainly concerned with life sciences. They recollected and made research on fossils, fungi, and plants. Their attitude may be new but their texts and practice, however, still reflected traditional classical scholarship. This academy was the natural evolution of princely courts that were fashionable in the Renaissance. This time, however, the establishment coincided with the beginnings of the Scientific Revolution, which granted it a privileged status. The invention of the telescope, reported to have taken place around 1608 (Applebaum, 2000, p. 634), and the microscope, only a few years after, were opening up new horizons to men of science but the change did not happen all of a

[6] The discovery of Proto-Indoeuropean and Grimm's Law in the eighteenth century and Verner's Law in the nineteenth can be said to be one of the most important discoveries in linguistics at that time and it shows to which point philology was also included in the Revolution.

sudden. The difference between humanists and scientists was not so sharp and multidisciplinarity was the trend at the moment. “Science for them covered a much larger field than it does now”, acknowledged Freedberg (2002, p. 9).

Another of the Italian academies taken as forerunners in the academy movement was the Accademia del Cimento, established in Florence by Duke Ferdinand and his brother Prince Leopold Medici, which stood active for ten years, between 1657 and 1667. The success of the Accademia del Cimento was partly due to the top-quality facilities made available to scientists. After all, one of the main reasons for the foundation of this academy was to make a public display of wealth of the Medici family. Leopold and Duke Ferdinand made sure they hired not only the best men of science but also the best instrument makers.

It has been repeatedly noted that the Accademia del Cimento set the model for the establishment of other scientific societies in the rest of Europe, especially the Royal Society of London (Applebaum, 2000; Feingold, 1989; Hall, 1954; Voss, 1980). Contemporaries already saw this link and in 1667 the astronomer Geminiano Montanari (as cited in Feingold, 1989, p. 230) referred to it as “supreme in Europe, heir to the tradition of Galileo, and a model to all academies north of the Alps, so much that in France, England, Holland and many other places experimental philosophy flourished in imitation of the Tuscan academy.” However, even if they have formal similarities, the heart of both institutions is different. The Cimento was the result of an expression of power displayed by Grand Duke Ferdinand and Prince Leopold. As Feingold (1989, p. 230) argued:

> The Cimento is perceived as an instrument of state, a propagandistic stunt staged by the Medici brothers in order to aggrandize the Tuscan Court by conveying the impression of its continued advocacy of the Galilean legacy. […] the Cimento was never granted a legal charter, or even a set of rules.

Unlike the Royal Society, there was no inner organization and membership was only based on the prince's decision. However, it would be also unfair to undermine the Cimento on the sole grounding that it lacked organization. The Medici brothers were indeed committed to science and they regularly attended and took part in debates. So, this Medici academy was, as the Accademia dei Lincei, a scientific organization situated between princely courts and scientific societies.

It is true that there are some important points in common between the Royal Society and the Accademia del Cimento. The origin of both societies was the result of private meetings and experiments. Similarly, the human element is also important here: the friendship between Giovanni Alfonso Borelli (1608-1679) and Carlo Rinaldini (1615-1679), promoters of the Cimento and Isaac Barrow, professor of geometry at Gresham College[7], has been verified. In fact, Barrow (1630-1677) is known to have traveled to Florence and have met the Medici brothers there. Feingold (1989, p. 236) noted this communication between Italian men of science and the rest of scholars in Europe and claimed that

> Italian savants were aware of the scientific activities of their English (and French) counterparts, and such an awareness facilitated the organization and proceedings of the Cimento – just as reports regarding its own activities stimulated, in turn,

[7] Gresham College was the first meeting place.

> the researches of members of the Royal Society and the Académie Royale des Sciences.

The point Feingold (1989) was trying to make here is that it is not possible to say that the Royal Society was modeled upon the Tuscan academy only because the latter is chronologically previous. This is rather a case of a bidirectional process in which both institutions influenced each other as a result of a general necessity of the time.

The foundation of the London Royal Society, an institution that was to become one of the main bastions of the new scientific method, was to take place in the early 1660s. Springing from a series of networking affinities among Cambridge and Oxford scholars, the history of the Royal Society is usually associated to two institutions. On the one hand, Gresham College, founded in 1598, and located in London was the place chosen for the first sessions of the Royal Society. The college, with a strong Puritan orientation, had six professors (rhetoric, divinity, music, physics, geometry, astronomy and law). It set a tradition for the practical application of scientific ideas in the solving of problems and the collaboration between artisans and scholars. On the other hand, Wadham College, located in Oxford was another common meeting place for scholars, who began to gather there reasonably regularly in the 1640s.

After a much-obliged interruption during the years of the Interregnum, the 1660s started with a renewed interest in the idea of establishing an academy of science. As a result, in 1662 the Royal Society of London for the promotion of Natural Knowledge received its first chapter. Its motto, *Nullius in verba*, declared the empiricist orientation of the Society. The activity in its early years was concentrated in weekly meetings where experiments were performed and witnessed by its members. The Fellows did not receive money from the State and the institution remained independent, also if the

findings and inventions were often used for state-service. The publication of the *Philosophical Transactions*, which began in 1666, was a complete success. The journal included discourses read by the Fellows and letters on a variety of scientific topics. The *Transactions* were translated into Latin and the Académie des Sciences launched a journal in imitation, called *Journal des Sçavans*.

France soon took over Britain in the establishment of academies related to the sciences and the arts. After Cardinal Richelieu had founded the Académie Française, an institution devoted to the study of the language, Colbert was one of the leading figures in the foundation of a scientific academy. The Académie Royale des Sciences was created in 1666 with the distinguishing feature that it was sponsored directly by the Crown. The French Académie differed from its British counterpart in its chore configuration: it had a hierarchical structure, with a small number of members that were state-appointed, received an income directly from the Crown and were expected to devote part of their time to work on issues relevant to the State. Originally there were two groups: mathematicians (astronomers and physicists) and philosophers (chemists, physicians and anatomists) They met twice a week at the Royal Library in Paris. In the eighteenth century the situation changed slightly, as more members were accepted and the configuration became more elaborate. Paris took the initiative but, little by little, the idea spread and small province towns started to settle up their academies of arts and sciences. In approximately forty years, the country was a ramified network of academies and scientific societies[8]. There was an important increase of not only national

[8] Apart from the Académie française (1653) and the Académie des sciences (1666), some of the scientific academies founded in this period are the Académie de peinture et sculpture (1648), the Académie des Inscriptions et Belles Lettres (1663; dedicated to History and Philology) and the Académie de musique (1672).

academies, but also secondary and tertiary societies, especially after 1750. As Coley (1991, p. 244) pointed out:

> The academies became the embodiment of the harnessing of rational knowledge, as against ignorance and superstition, both for its own sake and for political reasons to do with the application of such knowledge to the improvement of economic life.

Germany took a slower rhythm. One of the first academies to be founded was the Societas Ereneutica, established in the Mecklenburg city of Rostock in 1652. German academies, however, were lagging behind Italian, British or French institutions. In 1700 the Kurfürstlich-Brandenburgische Societät der Wissenschaften was founded in Berlin and it would soon catch up with other European academies. Leibniz, the founder of the Societät envisaged scientific academies as a utilitarian enterprise, focusing on the social and economic uses and criticizing those who considered science an entertainment. To keep up with this practical spirit, some changes were introduced at the end of the century to satisfy the necessities of the humanities. Consequently, the Societät came to be made up of four sections, namely experimental philosophy (natural history), mathematics, philosophy and belles lettres (history, languages and literature).

The academicist enthusiasm reached all countries in Europe and every state made sure to have an academy of science founded. Hence, the Real Academia Española (Royal Spanish Academy) was founded in 1714 and the Российская Академия Наук (Russian Academy of Sciences) was established in Moscow in 1724. In 1731, the Royal Dublin Society was founded and eight years later the foundation of the Kongliga

Vetenskapsakademien, the Royal Swedish Academy of Sciences took place[9]. At the end of the eighteenth century, Europe was organized under a long series of national and regional academies. Communication among academies was frequent but, in spite of that, no supranational institution was founded to try to organize the efforts of all the members and make them proceed in one single direction. One plausible reason may be that scientific societies remained mainly a state business. Voss (1980, p. 55) studied the relationship between academies and states and concluded that there was an apparent proclivity from monarchic states to organize science in academies: "Ingesamt gesehen kann man sagen, dass die Akademie die typische Organisation der Wissenschaftspflege in monarchischen Staaten des 18. Jahrhunderts darstellt. Republikanische Staatskörper wie die Schweiz und die Niederlande spielten in der Akademiebewegung keine große Rolle." Taken with perspective, this is a logical historical consequence as academies evolved from a special kind of princely courts. It is easy to expect, then, that the development of academies was easier in monarchic countries.

The proliferation of scientific academies also resulted in the development of parallel institutions, necessary for the practice of science, such as libraries, archives, archeological collections, botanic gardens, laboratories, publishing houses and observatories, among others. This, of course, fostered the specialization and professionalization of science that was to take place in the nineteenth century. In the last years of the seventeenth century experimentation and verification had replaced mere accumulation of facts, which had called the attention of scientists of the previous generation such as, for instance, those working at the Accademia dei Lincei. Likewise,

[9] For more information about the foundation of national academies of science, the webpage of the "Scholarly Societies Project", a project sponsored by the University of Waterloo Library offers an extensive account of the history of scientific institutions.

the initial impetus in propaganda for the new philosophy faded away and was replaced by discussion of the works produced by scientists. The level of abstraction and technicality augmented. Hill (1965, p. 200) noted that

> it was found, for instance, that as scientific books became more truly technical, more fully devoted to describing research (rather than useful textbooks or practical manuals), the publishing trade refused to handle them unless large sums were laid down.

This could be taken as a piece of evidence showing that the Scientific Revolution had already permeated language and that neologisms and technical vocabulary were common in scientific register.

Apart from the Royal Society of London, other English-speaking nations also took part in the institutionalization movement during the Scientific Revolution. Concerning Scotland, Bartholomew (1991) pointed put that the "Act of Union" (1707) meant a development of Scottish science. Culture and science were the response to the loss of political sovereignty. He named the chemist William Cullen (1710-1790), the philosopher David Hume (1711-1776), the social philosopher Adam Smith (1723-1790), the geologist James Hutton (1726-1797), and the physician Joseph Black (1728-1799) as the leading figures in Scottish science.

In the case of North American science, the backwardness of scientists and institutions in the seventeenth and eighteenth centuries has been the most highlighted issue (Greene, 1984; Stearns, 1970). The colonial and the early national periods had not been very successful in the history of North American science and the level and maturity of American philosophers in those years did not reach that of their European

colleagues. North American intellectuals were concerned with the creation of a nation and a state and it took time until an institutional base for the development of the sciences was established. As a result, it was not until 1780 that North American scientists started to make insightful contributions to Western science (Greene, 1984). Embedded in the American ideals of progress, the new science and its search for practical applications and solutions for specific problems was a success. In consequence, agriculture was as important for U.S. scientists as physics, astronomy, mathematics or mechanics.

U.S. scientists were, at the beginning of the nineteenth century, still largely dependent from Europe. Leading figures were usually educated in Edinburgh or London. In spite of the lack of political agreement between the two countries, U.S. scientists preferred a pro-British understanding of science and usually rejected French Cartesianism. All the political hostility seems to have evaporated in the scientific field and the U.S. did not cut its umbilical cord until more than one hundred years after its independence. Lack of equipment and facilities may be one of the causes for this dependency. As a newborn country strong universities, libraries, museums and publishing houses that could facilitate the work of scholars were lacking. Similarly, communication between states was slow, especially considering the size of the country. Patronage was scarce as well: the government had other priorities in mind and a wealthy leisured class that could serve as a patron simply did not exist. Popular science was not one of America's strong point either. “The flowering of popular science requires not only a potential audience but also some degree of maturity in science itself as well as a sense of security and leisure on the part of the scientist”, indicated Greene (1984, p. 21). Scientific societies, such as the American Philosophical Society and the American

Academy of Arts and Sciences published their *Transactions*. The scarcity of publications and the short life in the issuing of journals, however, may be taken as an indicator of the lack of audience of specialized scientific publications.

Philadelphia was the city chosen to become the capital of the sciences in America. Step by step, all the facilities necessary for modern science were built here and the city outpaced either Washington, D.C. or Boston. The American Philosophical Society grew in Philadelphia thanks to the support of its president, Benjamin Franklin. The life of the society was not as lively as that of the Paris and London academies. Members held gatherings but these were took place once or twice per month. The Society also had a journal in which its *Transactions* were published. The publication was, as in the case of the meetings, rather vague and in forty years there were only seven volumes published in intervals of approximately five or six years. Philadelphia also became famous for the development of the natural sciences. Its botanical gardens made the city the center of natural history in the country, as well as a tourist attraction. Equally important, Peale's museum, inaugurated in the late 1780s offered a collection of natural history-related items, which is a remarkably early date especially if we take into account that the foundation of the Academy of Natural Sciences of Philadelphia took place in 1812.

Even if Philadelphia was the main city to excel at the scientific level, the Boston area grew strong in creating a tightly unified network of associations and scientific institutions. The American Academy of Arts and Sciences was established there in 1780 by John Adams. Unlike the American Philosophical Society, the Boston Academy did not have a regular meeting place and members could only gather four times per year. Other satellite institutions near Boston were the Massachusetts Society for Promoting Agriculture (1785) and the Massachusetts Historical Society (1791). Even if other cities

and universities like Philadelphia or New Haven –the city in which Yale university is based– stood up to the level of excellence of New England, facilities and the long-standing university tradition made Boston recover all the prestige it had enjoyed in colonial times.

The city of New York, the real commercial capital of the country, had more difficulties in settling down a tradition for learning and science. Columbia College had been established in colonial times but it could not compete with the universities in New England. Perhaps more important was the development of the Military Academy at West Point in the 1820s and 1830s. Following the utilitarian and applied understanding of science characteristic of American science, the Academy provided instruction in engineering and related sciences. The evolution of science in other states followed a difficult path. Communication with Western states was not an easy task and the states in the south of Maryland remained rural and agricultural. The contrast between the urbanized, cultural north and the agricultural plantation-based south has always been an issue in the history of the country and it definitely affected the evolution of science.

It is unquestionable that the academy movement brought a series of improvements to science. Among the main benefits of this movement we could cite the encouragement it meant for scientists. Academies contributed to the professionalization of science and they gave a social role to the figure of the scientist. Talking strictly about science, the establishment of academies all over Europe increased collaboration between scientists, introduced cross-examination and extensive review to scientific work and fostered research. Equally important, it served as a consolidator of the Scientific Revolution.

Stewart (1992, 2008) studied the role that entertainment played in the new science and he successfully related it to economic considerations. One of his basic claims is that

modern philosophers pursued science also for entertainment's sake. He noted the example of John Theophilus Desaguliers (1683-1744), who was rebuked in 1725 by Newton because of the scarcity of experiments he had presented to the Royal Society. "Desaguliers' undiplomatic reply", Stewart (2008, p. 11) reported, "was to offer demonstrations of experiments he had already happily provided to paying audiences at his public lectures." The economic aspect seems incipient here and, in present-day speech, we would be talking of show business rather than science for its own sake. Utilitarianism was, after all, one of the main tenets of the new science.

The trade of scientific instruments was also another practical consequence of the Scientific Revolution. Sales in telescopes, microscopes and other scientific gadgets created an international commercial network. Not only buyers and manufacturers but also commercial representatives and even spies were involved in this new industry. The economic component is revealed, thus, as crucial in the success of the new philosophy. The new science needed funding and a reconciling attitude with other existing learned institutions would not, thus, satisfy their necessities. It was more intelligent to raise awareness by exaggerating the originality of approach than to present it as a logical consequence of long-established patterns. Roberts (1991, p. 234) commented on membership to the Royal Society:

> There was a predominance, throughout the period to 1700, of members from the professional and landed classes, as well as from the government and the court. To a certain extent, the founders deliberately cultivated an aristocratic strand in the membership in order to give the new institution and the new learning which it wished to promote a high standing, commensurate with the status of the royal patronage that it hoped to secure.

This quotation explains the fact that members of the Royal Society deliberately decided to differentiate themselves from universities. Scholars had to show they were different and more useful to attract the attention of the aristocracy, who, unlike universities, could provide them with patronage. The publication of the *Philosophical Transactions* gave the external visibility that the Society was longing for. Similarly, the publication of Sprat's *The History of the Royal-Society of London, for the Improving of Natural Knowledge* only five years after the Society's official foundation can also respond to this eagerness to claim visibility.

Feingold (1989) pointed out the bad reputation that universities have merited in the history if science. Normally, universities have been described as authoritarian and reluctant to the implementation of new methods and findings. Their role in the scientific revolution is thought to be especially negative. Quoting, E. Asby, Feingold (1989, p. 31) made it clear that universities had very little to do with the Scientific Revolution:

> The history of this Scientific Revolution lies almost completely outside the universities. It is true that Harvey did his best work in Padua, and Newton taught at Cambridge. But these circumstances were incidental to their discoveries. On no sense can universities of Europe be regarded as instigators of the Scientific Revolution.

The main clash between universities and the new scientific method was associated with the approach to science. Whereas universities tended to stress the importance of book-learning, the new academies of science approached knowledge through experimentation. *Nullius in Verba*, the motto of the Royal Society could thus be seen as an epitaph to book tradition. In England the situation was critical for universities. With a relatively short history of local universities, science under Queen Elizabeth I's reign was

primarily, as in other Germanic countries, a merchant's activity. It was easier then for followers of the new science to depose universities. They wanted to prove that their new method and their institutions were filling up an existing vacuum. They had to fight with a long-established university tradition and propaganda was one of their main assets. This position soon met with criticism from other scientists who saw the benefits of the new science but did not believe that it was incompatible with university knowledge. Thus, the reputed mathematician John Wallis (as cited in Feingold, 1989, p. 34) corrected the anti-university discourse of Henry Oldenburg, Secretary of the Royal Society and responsible for its foreign correspondence:

> I would not have you insist on that argument that the University doth not meddle with Experimentall Philosophy. For it is a great mistake, (Experimentall Philosophy being as properly appertaining to the Constitution as any other...) You should rather say it is no dispargement to the Universities, for others to pursue philosophicall studies also.

Feingold (1989) presented a series of evidence showing that universities were not against the Scientific Revolution. Professors then were aware of the change in method and they combined Aristotelian tradition with the use of new laboratories and instruments. They even encouraged students to make experiments even if universities lacked the money to acquire all the equipment needed for avant-garde practices. According to Feingold (1989) the discredit of universities during the Scientific Revolution shows to what extent academicians succeeded in managing the control of the panorama of science by devaluating the work carried out by other contemporary scientists. Then, it could be claimed that the present-day idea of a harsh opposition between universities and academies is partly a result of discourse. There were indeed

many scholars who did not understand the new philosophy as incompatible with university tradition. The Scientific Revolution implied, no doubt, a change but it also encompassed old traditions. Many scholars have remarked that experimental philosophy had ailed to get rid of old, obscure practices –old and obscure according to our present-day perspective–, such as alchemy, astrology and magic. Why should not it include old Aristotelian and university traditions as well? Some would say that rejection of the latter is indeed the reason why it was called a revolution, because it included a rebellious, nonconformist component. I would rather argue that, apart from being excellent scientists, experimental philosophers knew how to control discourse and sell what they were doing.

Apart from the indisputable fact that academies took control of the science panorama during the seventeenth and eighteenth centuries and settled a model that is still followed today, it is possible to approach this phenomenon in a critical way. By creating a standardized powerful monopoly, some marginal groups, especially women were excluded from the mainstream movement, as it will be discussed in the next section.

1.1.3. Women scientists and science outside the academia

As it is well known, the situation for women scientists –or any woman interested in pursuing learned activities– could be described as harsh and difficult until the twentieth

century[10]. In the case of this study, the low rate of texts written by women makes it even difficult to reach valid conclusions about the use of nominalizations according to sex. However, any attempt to outline the panorama of science in early modern and modern Europe and North America without presenting the situation of women would be incomplete. Indeed, it would be quite inaccurate to judge women's involvement in science only by the number of publications they made. As Schiebinger (1995) showed, the notion that women did not take an active role in science until the twentieth century is a wrong idea that originated in the nineteenth century. In early Modern and Modern Europe, women had a more active role in science but it is also true that their presence in scientific circles was not as ubiquitous as that of men's, women of that time could resort to non-canonical ways to access knowledge and practice science. Scholars (Burke, 2000; Schiebinger, 1995) have emphasized the importance of prince courts in the Renaissance together with salons in the seventeenth and eighteenth centuries as alternative places of knowledge with relatively easy access for women. It is a verifiable fact that scientific works written and published by women are few. The real implication of women in science, however, remains very difficult to demonstrate without reliable written materials to prove it.

Quoting Virginia Wolf (1929), a woman with learned aspirations would rather forget about them because society then did not allow the existence of educated women being displayed in public. In *A Room of one's own*, Wolf (1929) presented an imaginary character called Judith Shakespeare, sister to the famous playwright, who, sharing with

[10] It was not until 1945 that the Royal Society accepted women as full members. The first two women to become fellows were Kathleen Lonsdale, a crystallographer (1903-1971), and Marjory Stephenson, a microbiologist (1885-1971). Previously, important women scientists had been named "Honorary" Members, as in the case of the famous astronomer Caroline Herschel, who was elected Honorary Member of the Royal Astronomical Society in 1835, when she was already 85 years old.

his brother a gift for poetry and aspirations as an author, travels to London and ends up committing suicide because of the complete lack of opportunities. Leaving the poetic flaw aside, that could have been the fate of the German astronomer Maria Winkelmann (1670-1720). This German astronomer lived in the eighteenth century. She married Gottfried Kirch, a respected member of the Berlin Academy of Sciences. In his papers, Kirch (as cited in Schiebinger, 1995, p. 85) acknowledged that Winkelmann was more than a mere assistant:

> Early in the morning (about 2:00 A.M.) the sky was clear and starry. Some nights before, I had observed a variable star, and my wife (as I slept) wanted to find and see it for herself. In so doing, she found a comet in the sky. At which time she woke me, and I found that it was indeed a comet... I was surprised that I had not seen it the night before.

As a result of team work, the couple found several stars. When Kirch died, Winkelmann applied for an assistant calendar maker position at the Berlin Academy of Sciences in order to be able to maintain her children and continue her scientific investigation. In spite of her credentials and the great support shown by some members of the academy, she was never elected. The example of Maria Winkelmann is not extraordinary and the issue of women's involvement in science was often an ignored but latent debate after the Scientific Revolution. Indeed, it was deeply rooted in the idea that women were not capable of abstract intellectual activities that had been prevalent from the Middle Ages.

1.1.3.1. Books for and about women: the consideration of women and intellectual activities

To fight against the omnipresent idea that men are superior to women both bodily and mentally, we find myriads of authors defending the equality of both sexes. During the Renaissance, Giovanni Bocaccio's *De Claris Mulieribus* (1355/1359), Christine de Pizan's *The Book of the City of Ladies* (1405) and Henricus Agrippa's *Female Pre-eminence* (1532) were important books written to vindicate the intellectual aptness of women. One important advocate for women was the Italian Baldassare Castiglione. His most important work, *Il Libro del Cortigiano* (1528), is divided in four books. In the "libro terzo" he focused in the desired qualities for a lady. In his understanding of a perfect courtier, men and women should portray different qualities. For him, a lady should be beautiful, cautious, dignified, modest and affable. Regarding education, women should show their knowledge: "uoglio che queſta Dona habbia notitia di lettere, di muſica, di pittura: & ſappia danzar, & feſteggiare" (libro terzo). Most importantly, Castiglione was a great defender of the equality of the sexes. He refuted those theories asserting that women were imperfect beings. For him, women were even intellectually superior to men:

> Se nell'animo, dico che tutte le coſe, che poſſono intendere gli homini, le medeſime poſſono intendere anchor le donne: & doue penetra l'intelletto dell'uno, pò penetrare etiandio quello dell'altra. […] non è dubbio che le donne, per eſſer piu molli di carne, ſono anchor piu atte della mente: & di ingegno piu accommodato alle ſpeculationi che glihomini.

Against those who claimed that women aspired to be men to reach perfection, Castiglione argued that women's interest in becoming men was actually due to the control exerted on them by men: “Le meſchine non deſiderano l'eſſer homo per farſi piu perfette, ma per hauer libertà, & fuggir quel dominio, che glihomini ſi hanno uendicato ſopra eſſe per ſua propria authorità” (libro terzo). Even if Castiglione is not directly concerned with women scientists, his book gives us a good insight on how women were considered at that time. This can serve also as a good indicator of the beginning of a change in mentality that was starting to take place in Europe. Castiglione's work was a bestseller and in 80 years there were more than 108 editions and translations into English, French, Spanish and Latin. This is a clear marker of the popular support of his ideas in early modern Europe.

In the seventeenth century we find great women defenders like Margaret Cavendish and Mary Astell. Lady Margaret Cavendish, duchess of Newcastle (1623-1673) is one of the only women in seventeenth-century England that dared to write openly to criticize women's exclusion from the sciences. Being a noble lady, she is also considered a natural philosopher. She did not receive a learned education but she was given a lady's education according to her rank. Cavendish' main argument was that women's subordination to men in society was due to a lack of opportunity, which was the result of inappropriate education. In the preface of her *Worlds Olio*, she acknowledged (as cited in Schiebinger, 1995, p.48) that in most cases women's way to knowledge was through men: “Most Scholars marry, and their heads are so full of their School Lectures, that they preach them over to their wives when they come home, so that [the wives] know as well as what was spoken, as if they had been there.” Margaret married William Cavendish and she entered the Newcastle circle, which provided her

with all the learning she could not access before. However, even within that circle, she suffered from isolation and was not well regarded by all scientists (Baker, 2002) but she discussed science mainly with members of her family. Living an accommodated country life with her husband, she wrote the *Philosophical Letters* (1664), the result of her systematic study of works of natural philosophy of her time. In this book, she criticized Hobbes' and Descartes' theories and preferred rational speculation to experimentation.

At the end of the seventeenth century, Mary Astell (1666-1731), a middle class woman, coming from a merchant family from Newcastle revolutionized English society. Astell received her education thanks to one of her uncles, who had studied at Cambridge and had no children of his own. Like Cavendish, she was very concerned with the education of women. In 1697 she published *A Serious Proposal to the Ladies*, which became a rapid success and her most important work. In this book, she called on women to widen her intellectual capacities and aspirations. To achieve this goal, she proposed the creation of an intellectual retreat, a secular convent in which women could cultivate their minds. In this book, Astell (1697) blamed men for the inferior position of women but she also acknowledged that women were in that situation for their own will, because they wanted to please men. The purpose of the book is to take women out of this unfair situation. Accordingly, she (1697, p. 14) exhorted: "Ladies, I deſire you wou'd aſpire, 'tis a noble and becoming Ambition, and to remove ſuch Obſtacles as lie in your way is the deſign of the paper." To fulfill her goal Astell (1697, p. 8) addressed directly to women in an assertive way: "How can you be content to be in the World like Tulips in a Garden, to make a fine ſhew and be good for nothing". Education was one of the main vindications that she claimed for; it was an obligation of parents to provide their children with a good education that would make them –both children and parents–

happier and have a better life. The heart of a right education for women would be to develop the skill of judgment. Women should not be obsessed with learning many foreign languages or reading big amounts of books; they should rather concentrate on good works and try to understand and get the most out of them. In the following quotation, Astell (1697, p. 21) explained what she expected from women's education:

> She who rightly underſtands wherein the perfection of her Nature conſiſts, will lay out her Thoughts and Induſtry in the acquiſition of ſuch Perfections: But ſhe who is kept ignorant of the matter, will take up with ſuch Objects as firſt offer themſelves, and bear any plauſible reſemblance to what ſhe deſires.

The idea of an academy for women caused a stir among her contemporaries and Astell received patronage from wealthy ladies –including Queen Anne– to help bring her idea to fruition. French writers were very prolific during the years of the Scientific Revolution in writing essays and treatises championing for the equality of women[11]. The importance of these writers is considerable if we take into account that until the eighteenth century anatomists did not undertake a real revision of old traditional opinions about women's anatomy and ability to science. Men and women were thought to be different and theories about humors had been replaced by theories claiming that women were imperfect, not-fully-developed men (Schiebinger, 1995).

[11] A brief account of French works from the seventeenth century dealing with this topic may be: Marie le Jars de Gournay's *Egalité des hommes et des femmes* (1622), Samuel Chapuzneau's *L'Académie des femmes* (1661), Jean de la Forge's *Le cercle des femmes sçavantes* (1663), Louis de Leslache's *Les avantages que les femmes peuvent recevoir de la philosophie et principalement de la morale* (1667), François Poulain de la Barre's *De l'éducation des dames pour la conduite de l'esprit dans les sciences et dans les moeurs* (1674), Bernard Le Bovier de Fontenelle's *Entretiens sur la pluralité des mondes* (1686) and Gilles Ménage's *Historia mulierum philosopharum* (1690).

One of the most important defenders of women in the seventeenth century was François Poulain de la Barre. This French philosopher studied at the University of Paris, where he received a training that stipulated that women were inferior to men. He then refused this scholastic antiwomanism and turned to Descartes' teachings about the aptness of both sexes for learned activities. He attributed women's lack of advance in the sciences to the fact that they employed more time in housewifery. He advocated for high education for women; for Poulain de la Barre, women and men were not equal but each sex could be socially useful for different tasks. Taking into account that at that time men and women were thought to be physically different, his ideas should be considered revolutionary. In his work *De l'Egalité des Deux Sexes* (1679) he made a thorough review of the scientific disciplines of the time and pointed out in what way women could be apt for those activities. He (1679) claimed that women could be good teachers, doctors, officials in the army or even dictators. This aptness for social tasks is partly derived from woman's discourse. His idea (1679, p. 40) was that women had a more direct style and their message was conveyed more clearly: "On diroit que se que les hommes ſe mettent dans la teſte en étudiant ne ſert qu'à boucher leur esprit, & à y porter la confusion. […] Les femmes au contraire, diſent nettement & avec ordre ce qu'elles ſçavent: les paroles ne leur coûtent rien." The issue of woman's discourse as opposed to man's was indeed an important topic in the seventeenth and eighteenth centuries and scholars of that time were conscious of those differences. Diderot (as cited in Schiebinger, 1995, p. 153) blamed this tendency to gallant poetic scientific style on the interaction with women:

> Women accustom us to discuss with charm and clearness the driest and thorniest subjects. We talk to them unceasingly: we wish them to listen; we are afraid of tiring or boring them. Hence we develop a particular method of explaining ourselves easily which passes from conversation into style.

The debate over men and women's discourse extended to scientific discourse itself. The seventeenth century saw the raise of the Scientific Revolution and a new empiricist method permeated science. For Schiebinger (1995, p. 151), the debate over scientific discourse hid a debate about ancient and modern science:

> In the seventeenth century there was the struggle between the ancients and moderns over the desired character of scientific language: should language retain the allegorical richness of the ancients, or adopt the more flat-footed precision of the moderns? In the eighteenth century scientists tried to cleanse "nature, the earth, the human soul, and the sciences of all poetry.

The debate was between old literary style and new precise scientific prose. Women's style was thought to be gallant, polite, aristocratic and poetic, as opposed to Bacon's virile and masculine style. Francis Bacon was an important figure in this debate and he was partly responsible for the misogynist flavor of scientific academies and institutions in England. Precursor of the scientific method, which would be adopted by the Royal Academy and all the leading scientists in England, Bacon had a masculine understanding of science. As Schiebinger (1995) showed, for Bacon calling something "masculine" was an appraisal, whereas calling it "feminine" or "affeminate" was an insult. For empiricists, man and science were active, they did things and they required energy and power. This idea became imbued in scholarly circles and it definitely affected very negatively women's involvement in science. This debate was indeed

relevant for scientists of that time and philosophers were aligned in two sides: those who would accept a peacefully relation between men and women and those who would declare a war of sex. Bacon and his followers were in the latter group (Eger, 1999).

In the eighteenth and nineteenth centuries, the number of publications defending the intellectual ability of women multiplied. In England, Mary Astell and Judith Drake published *An Essay in Defence of the Female sex: In a Letter to a Lady in 1721*. Following the high interest showed by French authors, German writers joined in the task of defending women. The work of anatomists in the second half of the eighteenth century led to the conclusion that there was no intrinsic difference between the nature of men and that of women. Men and women were considered "perfect in their difference" (Schiebinger, 1995, p. 191), each of them displaying their own characteristic features –physical and intellectual strength for the man and motherhood for the woman. This new configuration could perfectly fit into Darwin's evolutionary theory but it failed to secure equality for women, since women were thought to be perfect but hierarchically inferior to men. Leaving these considerations apart, it may be noted that this resexualization of the body meant an even further increase of essays and treatises defending women[12]. As we can see in titles, a slight shift of interest can be noticed. Scientific essays in the nineteenth century are not concerned with a simple defense of women, but they attempt to compare men to women and provide a biological

[12] Some examples could be Jacques-Louis Moreau de la Sarthe's *Histoire naturelle de la femme* (1803), Jouard, Gabriel's *Nouvel Essai sur la femme considérée comparativement à l'homme* (1804), Alphonse de Candolle's *Histoire des sciences et des savants depuis deux siècles* (1885) and Alphonse Rebière's *Les femmes dans la science* (1897). Among German studies we could highlight Amalia Holst's *Über die Bestimmung der Weibes zur höhern Geitesbildung* (1802), Johann Ziegenbein's *Aussprüche über weibliche Natur, weibliche Bestimmung, Erziehung und Bildung* (1808), Karoline von Woltmann's *Über Natur, Bestimmung, Tugend und Bildung der Frauen* (1826), J.J. Sachs' *Ärztliches Gemälde des weiblichen Lebens im gesunden und krankhaften Zustande aus physiologischen, intellektuellen und moralischen Standpunkte* (1830), Hedwig Dohm's *Die wissenschaftliche Emancipation der Frauen* (1874) and Elise Oelsner's *Die Leistungen der deutschen Frau in der letzten vierhundert Jahren auf wissenschaftlichen Gebiete* (1894).

explanation to prove that women could indeed undertake intellectual activities very successfully.

Apart from all the pieces of work defending women, another indicative point providing a trustful piece of evidence for the active involvement of women in science is the big number of scientific works aimed at women. In England there were journals aimed at the feminine audience like *The Athenian Mercury* (1690-1697), *The Free-Thinker* (1718-1721), and *The Female Spectator* (1744-146). Perhaps one of the most important scientific journals aimed at women was *The Ladies' Diary*, which was regularly published from 1704 to 1841. It contained almanacs, enigmas, mathematical questions and answers, quests, chronologies and remarkable events of the year, birthdays of the members of the royal family and main kings in Europe. The title of the journal, *The Ladies' Diary, Containing New Improvements in Arts and Sciences, and many Entertaining Particulars: Designed for the Use and Diversion of the Fair Sex* made an explicit allusion to women as intended audience. Several studies, however, have proved that, especially at the end of the eighteenth and beginning of the nineteenth centuries, many men were assiduous readers of the journal.

In most cases, titles already contained the specification that the work was directed at women. Examples[13] of this may be John Harris' *Astronomical dialogues between a gentleman and a lady* (1719), Jasper Charlton's *The Ladies Astronomy and Chronology in Four Parts* (1735), James Ferguson's *Easy Introduction to Astronomy for Gentlemen and Ladies* (1768) and Denison Olmsted's *Letters on Astronomy, Addressed to a Lady in which the Elements of the Science are Familiarly Explained in Connexion with its Literary History* (1841). Now the question remains: if we can find pieces of evidence

[13] Samples of all these texts are contained in *CETA*.

showing that women in the seventeenth and eighteenth centuries were urged to engage in intellectual activities, how is it that we do not find socially recognized women scientists then? The issue was not the lack of interest from women, but rather an enduring and systematic refusal from scientific institutions. The next section is concerned with a short history of this refusal.

1.1.3.2. Places of knowledge, the figure of women scientists and their exclusion

As discussed in previous sections, the idea that universities and academies as the only places to make science originated in the eighteenth century and it could be listed as one of the consequences of the Scientific Revolution that took place in the seventeenth century. Academies and universities were closed to women until the beginning of the twentieth century. Before academies took the monopoly of science, there were other places of knowledge that provided an easier access to women. Similarly according to Schiebinger (1995) the history of women scientists in England has been especially harsh in comparison to other European countries due to a series of unfortunate coincidences. Schiebinger (1995) analyzed the history of knowledge from the Middle Ages to the Enlightenment and established three different periods according to places where science was being made, namely universities, prince courts, science academies and aristocratic salons.

First, in the late Middle Ages universities took control of the access and practice of science. Universities were at this time closed to women and it was virtually impossible that a woman could be granted access[14]. Access to libraries, instruments and knowledge was strictly controlled. However, clerical life was an option also for men and some women –especially belonging to an aristocratic family– could pursue a life of praying and learning in monasteries[15]. The raise of universities and the Church of England, which meant the closing of many monasteries, can be seen as a displacement of women from the centers of knowledge. It may be noted, however, that university policies were not the same in all countries. The hermetic rejection to women in English universities was completely opposite to Italian universities, as there are examples of Italian female doctors[16] (Schiebinger, 1995).

The Renaissance brought a revolution in the approach to science and that resulted in a change in the socializing habits of science and, consequently, in the places where science took place as well. Attracted by the patronages offered by princes and kings, scientists progressively gathered in prince courts. Well-born women could have access to these courts and took part in the revival of learning. Writing and learning were not a skill required for members of the aristocracy in the Middle Ages. The Renaissance fostered the literacy rate among the privileged classes and women were somehow included in this revolution. The role of women in these prince courts cannot be clearly

[14] The degree of antiwomanism in universities has always been great. Women were thought to be a distraction for professors and consequently the universities of Cambridge and Oxford required celibacy to their members and professors were not allowed to marry until the end of the nineteenth century.

[15] Schiebinger cited St Radegung, abbess of Poitiers and Hildegard von Bingen as examples of literate religious medieval women.

[16] There were five women who received their doctorates in Italy: Bettisia Gozzadini (thirteenth century, Bologna), Elena Cornaro Piscopia (sixteenth century, Padua), Laura Bassi (eighteenth century, professor of physics at the university of Bologna), Maria Agnessi and Anna Manzolini.

established, given that there are no direct written records of the activities carried out there.

The importance of prince courts started to decline as the Scientific Revolution started and scientific academies were founded. Having both universities and prince courts as precursors, academies could have been more open to women. There are many clues that point out that the issue of admitting women in the French Académie des Sciences was really important. In its creation, the Royal Society aimed at bringing together knowledge from any origin. However, in practice the Society remained almost closed to working classes and to women. Margaret Cavendish seems to have been attended a session of the Society and she was even considered for whole membership but the idea did not succeed and no woman became a member until 1945. Following its feminist tradition, women were indeed accepted in Italian academies of science. Then, in the eighteenth century, another new place of knowledge, the salon, came into coexistence with academies. In Paris salons hosted by women, where academicians gathered, were fashionable. In these semiformal meetings, scientists, other women and young men eager to make a career in science gathered, rich women found young scientists to offer patronage to and science was discussed. In England, the influence of the Bluestocking circle was big (Egers, 1999; Harcstark-Myers, 1990). The Bluestocking circle was a series of informal gatherings of upper class and professional middle class men and women in the London homes of well-to-do society women. Lady Elizabeth Montagu, Frances Boscawen and Elizabeth Vesey were some of the ladies that started to host these gatherings in the 1750s. Apart from gatherings in London, the activity of the circle also extended to countryside visits and correspondence. All the members of this circle pursued a social and political modernization of society. Their

idea was not that of an abrupt rebellion but rather an effective modification of customs based on moderation. It can be said that the horizons of expectation for women scientists opened up in these salons and literate circles. Schiebinger (1995, p. 36) described the situation of women scientists in the eighteenth century as follows:

> Exclusion from academies, while it distanced women from the centers of scientific endeavor, did not end their participation in science. In the seventeenth and eighteenth centuries […] there were a number of women working in natural history and natural philosophy, as well as the experimental sciences. Though few in number, women made real contributions. It is important to understand how these and other women, though barred from universities and scientific societies, could nonetheless acquire the training required for work in the sciences.

This promising panorama for women did not change gear and it progressively died off under the all-powerful influence of misogynist universities and academies. The professionalization of science in the nineteenth century implied that universities became the only centers of science. The figure of the scientist became identified with that of the university professor and, having that path officially cut down, women were excluded from science.

Scholars have shown that the idea that women in the eighteenth and nineteenth centuries were not engaged in science is not accurate (Levin et al., 2000; Rossiter, 1982; Schiebinger, 1995). It would be more appropriate to claim that they were systematically barred from official science, that is, universities and scientific societies. The exclusion of women in science as rightful members was not debatable until the end of the nineteenth century and the beginning of the twentieth. However, there is evidence that women were actively involved in science. Traditionally, it has been claimed that women

were devoted to applied and technical branches of science; they were helpers, always in the background. This is, however, a mistake, as science was different then. The differentiation between formal and applied scientific knowledge did not originate until the nineteenth century. It may not be accurate to apply present-day notions of science to the study of the seventeenth and eighteenth centuries. In those times, the difference between formal and applied sciences was not developed. Unlike today, where practically the totality of recognized scientists work in universities, it was more common in those centuries to work for a prince or outside universities, which made it possible for women to access knowledge and science outside university circles. Galileo and Descartes, for instance, worked at prince courts. I will argue that the non-recognition of women is mainly related to three issues, namely social pressure and lack of opportunities, social class and the involvement in activities not considered science.

Concerning social pressure, philosophers and scholars enunciated theories that proved scientifically that women should restrain from intellectual activities and stay at home taking care of the children and the household. In the eighteenth century the Enlightenment cut off theories about women being imperfect men. Following the ideal that all human beings are equal, women went up one step in the social ladder. In the late eighteenth and early nineteenth centuries the woman question was resolved with the theory of sexual complementarity, that is, men and women were not superior or inferior; they were opposites and could not be compared. The theory of sexual complementarity "was designed to keep women out of competition with men in the public sphere and, at the same time, to preserve the family within the state" (Schiebinger, 1995, p. 224). Social theorists like Rousseau emphasized women's role in society and they claimed that woman's constitution determined her place in the physical and moral order.

Consequently, any intellectual aspiration from a woman was negatively regarded. Women's place is society was that of nurturers of the family and all their efforts should be directed at that goal. Any activity that could deviate women's attention from her social duties was not admitted. Science and personal development fell in that category, so any women with intellectual aspirations were regarded as treacherous deserters to her duties. This situation makes it even more difficult to track women scientists because in many cases, to avoid social pressure and criticism, women gave up social recognition and published their works anonymously, under the name of a member of the family –as in the case of Caroline Herschel– or under a pseudonym.

This situation was slightly different in the case of women of a high social rank. The social class issue clearly worked in women's favor. In a heavily hierarchical society, birth and aristocracy played sometimes a better role than gender. Female aristocrats were higher in rank than mere scientists. Of course, this configuration excludes from science women of a low birth interested in science. Schiebinger (1995, p. 65) explained this situation as belongingness to elite groups with access to knowledge:

> In the seventeenth and eighteenth centuries, natural philosophy remained a part of elite literary culture. Noblewomen were able to insinuate themselves into networks of learned men by exchanging patronage or public recognition for tutoring from men of lesser rank but of intellectual stature.

Noble women were also free from the role of nurturers. Either engaged in intellectual activities or in mere social entertainments, noble women usually had wet nurses and housekeepers to take care of the household and the children. As a result, those women with a predisposition for science had a relatively easier access and could

devote themselves to it. All the ladies from the Bluestocking circle were aristocrats or had a high social rank. Perhaps one of the most important scientist noble women was Lady Margaret Cavendish, duchess of Newcastle (1623-1673). Before the creation of scientific academies, many of these noblewomen attended regularly and hosted social gatherings with the leading scientists of their time. Learning and reading even began to be forms of leisure for the aristocracy –asking and answering questions, entertaining people in the court. This made countesses and duchesses get interested in learning, because they wanted to be good hosts and entertain their guests. The situation was not very promising, as ladies were usually reduced to the position of asking questions (Schiebinger, 1995, p. 19) or becoming patrons for young scientists in search of a living. Even indirectly, women were indeed involved in science.

But that was not always the case, apart from patrons and hosts, women did play an important role in science. We find, however, that there were many women involved in science-related activities that did not enjoy the form of address of scientists. One of the possibilities was that of being translators and commentators of texts. The work of women in this field was not undermined and many regarded this activity as feasible to be carried out by women. The job of translator was rarely considered a permanent occupation for women but it could guarantee them some money. Translators usually had –and still do– a very good command of a number of classical and modern languages. Most frequently women translators were concerned with journalistic and literature texts. That is the case of the English Elizabeth Craven (1750-1828), the Danish Dorothea Briehl (1731-1788), the German Amalia von Helvig (1776-1831), the Italian Maria Ardinghelli (1730-1825) and the Swedish Catharina Ahlgreen (1734-1800) and Anna Maria Lenngren (1754-1817), to name only a few. In many cases, translation also

involves specialized notions of a certain scientific discipline. Thus, translators must be acquainted with the situation of the discipline. Elizabeth Carter was famous for her translations of Epitectus but she was also responsible for the translation and introduction in England of Algarotti's *Newtonianismo per le Dame*, a highly technical text.

In other cases women acted as what has been called "invisible assistants" (Schiebinger, 1995), that is, women that worked side by side with men, mastered techniques and even made discoveries, but they were never socially recognized for their work. There are a good number of women astronomers that fall into these categories, such as Maria Winkelmann and Caroline Herschel.

The term *invisible* should be carefully used. Even if it is true that these women never got a job in any academy, university or scientific institution, it cannot be said that their job was unknown to the scientific community. Scientists in the eighteenth century did indeed know about women *assistants* and not everybody showed repulse to them, as can be seen in the following quotation, uttered by Jean de la Bruyère (as cited in Schiebinger, 1995, p. 23), one French Academic:

> I have not forgotten, gentlemen, that one of the principal statutes of that illustrious body of advocates admitting only those whom one judges the most distinguished. You will not find it therefore strange that I give my vote to Monsieur Dacier, though all the same I prefer Madame, his wife, if you would admit among you persons of her sex.

Another science-related occupation of women was that of midwives and nurses. These two disciplines are today considered scientific but that was not the case three hundred years ago. During many centuries, women had the monopoly in the theory and

practice of midwifery and nursing activities. Nursery was not established as a rightful discipline until the second half of the nineteenth century thanks to the work of the English nurse Florence Nightingale (1820-1910), who established the role and education of nurses-to-be after her experience in the Crimean War. However, long before that, women wrote specialized documents trying to explain midwifery and nursing activities which, on the light of our present-day perspective, could be considered scientific.

Concerning the education of women scientists, no single pattern can be traced. Even if it is clear that women scientists belonged mainly to high or middle-high classes, the impossibility of access to regularized higher education makes it difficult to create a trustful parameter to describe their education. In many cases, girls were educated at home by their fathers. One illustrative example is that of Sir Thomas More's (1478-1535) household. More believed in the benefits of a learned instruction, which held that an educated woman would be more pious, charitable and humble. As a result, he hired instructors for all his sons and daughters, who received an extensive humanist education and were taught Greek, Latin, grammar, rhetoric, logic, theology, philosophy, astronomy and medicine –that is, trivium and quadrivium disciplines–. As a result, two of his daughters, Margaret Giggs Clement (1505-1570) and Margaret More Roper (ca. 1505-1544) became translators of classic books (Levin et al., 2000).

In most cases, the refusal to access knowledge came from the mothers themselves, who might have been aware of the necessity of providing their daughters with an appropriate education and discourage them from fields not aimed at them. This was the example of Caroline Herschel. In the following extract, taken from a book written by

Mathilda Betham-Edwards in 1880, Herschel (as cited in Betham-Edwards, 1880, p. 96) described her mother's reluctance to let her husband give her a learned education:

> My father," she says, "wished to give me something like a polished education, but my mother was particularly determined that it should be a rough, but at the same time a useful one; and nothing further she thought was necessary but to send me two or three months to a sempstress to be taught to make and mend household linen. All that my father could do was to indulge me (and please himself) sometimes with a short lesson on the violin, when my mother was either in good humour or out of the way.

However, Herschel's eagerness to learn was great and she resorted to unusual ways of accessing knowledge. She was never accepted in any higher education institution but she taught herself all the basics of her discipline and became one leading scientist in it. As Betham-Edwards (1880, p. 124) mentioned: "the mathematical knowledge needed in her calculations she had to gather at odd times, chiefly during meals, when her brother could be freely interrogated." Family was indeed one of the most important support and without the help of fathers, brothers, husbands, uncles and cousins, the already-small number of women scientists would have dropped even further because in many cases male relatives were the only way of accessing knowledge.

University, the main exponent of science today, and science itself are still elite entities. Scholars have showed that seventeenth and eighteenth century incipient institutions proclaimed to be universal but failed in the realization. In this section it has been shown how women were left out of this movement and how the specialization and professionalization of science progressively restricted the number of rightful scientists.

Scholars may have pointed out those failures of the system but, in spite of all that academic effort, the system has not been modified. If any European or American university would reject a woman today, media would soon report on it and a public scandal would take place. However, we seem to have retained other kinds of taboos, such as access to education disregarding monetary concerns. University is, unlike it was in the eighteenth century, the only way of accessing science today. To become a professional scientist, one must acquire scientific knowledge in a university, preferably a reputed one. Self-taught scientists, disregarding their brilliance, are not likely to get a job anywhere. Modern Europe people may have been wrong in their thoughts about women but they were more open regarding access to science.

1.2. The discipline of astronomy

Astronomy, the scientific study of celestial objects, is one of the oldest and most popular sciences. In the traditional understanding of natural philosophy as a comprehensive matrix of knowledge, astronomy occupied a central position, it was "the archetype of exact sciences" (Camiña, 2013, p. 41), thanks to its methodology, which transformed data obtained from observation into computable data by the application of mathematics. The study of the heavens attracted the attention of ancient civilizations but it is not until the invention of the telescope that modern astronomy began. In fact, astronomy has been considered one of the disciplines where the effects of the Scientific Revolution are most remarkably seen (Butterfield, 1965). The fifteenth and sixteenth

centuries meant a total revolution for the study of astronomy as a science. In 1453 Nicholas Copernicus' reassertion of the heliocentric theory provoked a revolution and meant an impulse for other authors to formulate new theories. Thus, Johannes Kepler (1571-1630) was able to formulate a series of laws of planetary motion that described the elliptical orbits of planets and Galileo Galilei (1564-1642) used the telescope to discover the phases of Venus and the four moons orbiting Jupiter. Tycho Brahe (1546-1601) was one of the most important astronomers in the sixteenth century. Even if he worked without a telescope, he was able to produce an accurate account of the position of the planets and the stars. In the seventeenth century, Sir Isaac Newton (1642-1727) formulated the laws of motion and gravitation meant the birth of modern astronomy and astrophysics. His *Philosophiae Naturalis Principia Mathematica* (1687) was considered central in astronomy until Einstein's theory of relativity.

The Age of Reason cast a new light on astronomy and it meant the establishment of an independent astronomical discipline. As Rothenberg (1985, p. 118) pointed out “the older form of science persisted until 1700 in the form of almanacs”. Most scholars were concerned with the study, application and development of the Newtonian theory. The universe began to be conceived as a clockwork-like mechanism and astronomers devoted themselves to the calculation and prediction of planetary orbits. The amelioration of telescopes and other instruments, together with the impact of Newton’s theories, made it possible for astronomers to make several very important discoveries. In 1718 Edmund Halley discovered the stellar motion. Until this time, the sun, the moon and the rest of the planets were called the fixed stars. Halley, however, noticed the change of position of stars Sirius, Procyon and Arcturus. From fixed stars, celestial objects began to be considered like bees flying in the space. Halley also made

predictions for comet's orbits. In 1781 Uranus was discovered by William Herschel. It had already been noticed before but it was thought to be a star. The discovery of Uranus gave extra encouragement to astronomers because it showed that not everything had been discovered and that the Greeks had also made mistakes. Two years later, Herschel made another transcendental discovery concerning the motion of the Sun. Ptolemy's followers believed that the earth was the motionless center of the universe, and modern astronomers did not contradict this view. Measuring the movement of the stars, Herschel came to the conclusion that the sun also moved. The question of what was the center of the universe remained unanswered.

The eighteenth century was also important because it meant the final separation between religion and astronomy. Up to this century, astronomy was closely related to religion and theology. The immutability of the universe was seen as a proof of the existence of God. Any discrepancy with the doctrine established by the Church was seen as an offense to the institution itself and created bitter social debates. That was the case with Copernicus and Galileo, whose theories were considered heretic. It was not until the Enlightenment that we can find a sharp distinction between religion and astronomy. For scientists of that age, reason and the application of scientific laws could explain how Nature worked. Religious wars and revolts played in favor of astronomy, as science was used to dismantle the mystical view of the world produced by Catholicism. Puritans became thus one of the main facilitators of astronomy. The new discoveries made by Galileo, Halley and Herschel also challenged old considerations and fostered new reinterpretations of religion: "The appearance of new stars (novae) and comets, and the strange behavior of variable stars, all indicated that the heavens could change" (Murdin, 1985, p. 3).

The development of the applied branches of science meant a progressive improvement of the size and effectiveness of telescopes. As a result of this mechanic progress, astronomers in the nineteenth century were concerned with the discovery and description of new stars and all the remaining planets of the solar system, as can be demonstrated by Galle's first description of Neptune in 1846[17].

Murdin (1985) studied the work of astronomers in the modern period from a sociological point of view. The overall image provided by her is that astronomers were rather solitary people, alienated from society and rarely building social networks outside their working and family circles. Murdin (1985) was also concerned with the education of astronomers. She claimed that many famous astronomers were taught at home in the seventeenth century. Very few institutions could provide a satisfactory education in mathematics. Many astronomers, then, acquired the skills necessary for their studies as seafaring and navigators. However, "as education involved trade," Murdin (1985, p. 43) asserted, "it became a political issue and could not be left to academics." This is not seen as a negative consequence because this utilitarian use of science proved to secure social interest. The interrelation between astronomers and navigators in the last part of the seventeenth century was endless and considered of essential importance, as we can see in Isaac Newton's words (as cited in Murdin, 1985, p. 49)

> If instead of sending the Observations of seamen to able Mathematicians at land, the land would send able Mathematicians to Sea, it would signify much more to the improvement of Navigation and safety of Men's lives and estates on that element.

[17] Pluto was discovered in 1930 by Tombaugh but in the light of recent astronomical debate it is no longer considered a planet.

Scholars have been especially critical with the role of universities and the encouragement of astronomy. British universities –notably Oxford and Cambridge– were particularly reluctant to the introduction of the new scientific method and, consequently, of astronomy. Other scholars, however, reconsidered the role of universities and have remarked that even if they were not the avant-garde centers of advancement then, universities did indeed accept the findings of modern astronomy and slowly incorporated them to their curricula, as can be seen in the attempt to set up an observatory over the gatehouse of Trinity College (Murdin, 1985, p. 40).

Also characteristic of this time was the high degree of interdisciplinarity derived from the fact that natural philosophers were indeed specialists in several related branches (Camiña, 2013): Newton worked also in chemistry and was a professor of mathematics at Cambridge; Christopher Wren was not only an astronomer but also an architect. As Murdin (1985, p. 6) stated, "astronomy claimed a high proportion of the time of scientists from all backgrounds. Among professionals, amateurs and paid workers, there were some who were involved in the most central and serious work of the time."

The corpus material for this study includes texts written in English by English-speaking authors. Among all the (former) British colonies, American astronomers were the ones that acquired their independence earlier. The *Declaration of Independence* of the American colonies in 1776 was not only political but also social and linguistic and, consequently, the effects of independence rapidly affected the development of science in America. Greene (1954, p. 339) described the involvement of American astronomers as follows:

> American astronomers contributed no great discoveries either empirical or theoretical, but they kept abreast of the latest developments, made and published useful observations, and propounded theories of their own to account for what they observed.

Several authors cited John Winthrop as the responsible for the introduction of the discipline in the United States in the seventeenth century. During the colonial period, colleges in the area of Boston monopolized the astronomical activity of the country. Until the independence, American scholars depended on European journals for publication of their findings. In 1771, the publication of the *Transactions of the American Philosophical Society* partly mended this situation. The nineteenth century was an age of institutionalization in U.S. astronomy (Rothenberg, 1981), as well as in Great Britain. The Cambridge Branch of the American Astronomical Society, the Peter's Astronomical Society and the American Astronomical Society were created in 1853, 1858 and 1883 respectively, aiming at providing U.S. astronomy with an international status.

Concerning the practice of science, astronomy was a tremendously popular discipline. "Certainly no science exerted a profounder influence on Western thought in this period than astronomy" claimed Greene (1954, p. 339). Indeed, astronomy always attired the attention of amateurs. The distinction between professional and amateur practice started to originate in the seventeenth century. However, accessibility remained open to all social classes, from aristocrats to middle class people or craftsworkers. Science did not require very specialized instruments yet and "on the whole, men who stayed working steadily at home produced the most useful results" (Murdin, 1985, p 28). Astronomers did not really fall into the prototypical image of a member of the

Royal Society –that of a wealthy upper class man. Important astronomers were not only professors teaching at university but also landlords, churchmen or tradespeople. Likewise, astrologers and almanac makers are to be considered specialists in applied astronomy. In his study about popular astronomy in England between 1750 and 1850, Inkster (1982, p. 122) asserted that "the range of appeal of astronomy was greater than that of any other science for this period, although by the later years geology and natural history were usurping its popular role". Similarly, Greene (1954), Yeomans (1977) and Rothenberg (1981, 1985) reported on the existence of an extensive amateur practice in the field of astronomy. From these claims, I have chosen astronomy as the basis for this study. Besides, apart from being extremely popular, the modernization of astronomy during the Scientific Revolution featured some of the most distinguishing features of the movement: separation from the Church, the adoption of an experimental method, an institutionalization and later professionalization of the activity and a utilitarian reinterpretation expressed in the development of applied branches. As stated in the introduction, such profound changes must have left a trace in language, as this is the main vehicle for communication. The next section aims at introducing the situation of English at the period and accounting for the establishment of a scientific register.

1.3. English in the modern period

Modern English covers a wide period of time, which is normally believed to have started around 1500 (Barber, 1976, 1993; Görlach, 1991). As Camiña (2013) remarked

the specific temporal boundary between early Modern and late Modern English is more evasive and some scholars pointed out at the end of the seventeenth century while others have delayed the date one century (Barber, 1976; Görlach, 1994; Moskowich, 2001). The end of Civil War (1651) and the transition from the Stuart to the Hanover House (1714) on one hand and the *Declaration of the American Independence* (1776) on the other have been cited as possible boundary dates. Independently of temporal concerns, Millward (1988) pointed out three of the most important events in history that helped shape modern English during the seventeenth, eighteenth and nineteenth centuries, which were the exploration and colonization of the American continent, the Industrial Revolution and the American Revolution. Among these events, the Industrial Revolution, initiated with James Watt's invention of the modern steam engine, was the most influential event, since it meant the introduction of massive technical vocabulary from Latin, Greek and other languages that fostered the development of a scientific register in English.

Since the seventeenth century English had a commonly accepted standard variety. Görlach (1999) attributed this to four main factors. On the one hand, the homogeneity of an educated variety settled in and around London contributed to the rapid spreading of a standard variety recognized by everyone. That standard variety was also backed up by the existence of a literary tradition in the language. On the other hand, the establishment of the variety of the Southeast of England as the written standard led to the stigmatization of regional dialects. Finally, the receding use of Latin and French in educated circles coincided with a resurgence in the use of English.

The situation of the language at the turn of the seventeenth century was increasingly expanded to cover new aspects of life. At that time, English was struggling

to become a respectable language for scientific use. Even if language was already a matter of national pride and a symbol of nationalism, Latin was still the prevalent language for writing science not only in England, but also throughout Europe. The eighteenth century was a time of consolidation for English. The language covered now all domains of life, either written or oral, and after the troubled sixteenth and seventeenth centuries, full of linguistic as well as religious and political changes and revolutions, the eighteenth century appeared as a time for consolidation. As Görlach (2001) pointed out, this was a time of extreme authoritarianism, normally referred as the 'Augustan Age'. In this century, Britain was a growing power. The battle of the Boyne (1690), the capture of Gibraltar (1704) and the *Act of Union* (1707) reasserted Britain's power in Europe, whereas the American colonies kept growing and paying their contributions to the country. Population on the eighteenth century remained stable and mainly rural. The Industrial Revolution provoked a migration to city areas at the end of the century but in any case, the figures of migration movements remained rather low until the nineteenth century.

One of the main advancements of this century was that printed materials became accessible to more people. A direct consequence of this was the rise of newspapers and journals and the specialization of printed publications for a certain audience. Patronage from wealthy aristocrats was still very common but it progressively died out, as the number of paid printed journals and newspapers multiplied. Grammars and dictionaries also proliferated in the eighteenth century. English was starting to be the language of general instruction and, consequently, teachers and students needed reference books. Several authors (Bailey, 1996; Görlach, 1999) signaled the important social changes that took place in the nineteenth century and affected the language. Among these changes,

perhaps the most important one was that the widespread of English and its standard written form awakened an unparalleled interest in grammar books and dictionaries by common speakers, which was especially remarkable in the second half of the nineteenth century.

Another very important social change was the intense increase in population. Thanks to colonization and to the multiplication of population in countries where English was already the native language, English became a world language. Except for Ireland, where the potato famine cropped the population, all English-speaking countries increased in population, and, consequently, from the 26 million people that had English as their mother tongue at the beginning of the century, we find a population of 126 million English speakers 100 years later. The colonization carried out by both the United Kingdom and the United States exported English worldwide and, in sociolinguistic terms, it meant an increase in bilingualism. As they were in a position of superiority, English speakers did not need to learn native languages but that was not the case with the population of colonized territories. Bilingualism, multilingualism, together with the creation of pidgins and creoles became, hence, direct linguistic consequences of British and American colonialism.

The extension of literacy is another well-known change that took place in this century. The nineteenth century started with very low numbers of literate population. There was literally no reward for literacy among working classes. Society was for the first half of the century rather oral. The population had contact with written texts and literacy rates were high but that does not imply that active reading was a common activity. Common people would normally get informed of news by oral accounts, they would attend public readings of the Bible and, even if they had basic reading skills, they

would not practice private reading very frequently. Bailey (1996) highlighted the abolition of the tax on paper in 1861 as one of the main reasons why newspapers started to proliferate. Concerning the use of newspapers in the first half of the century he (1996, p. 26) commented:

> In 1792 newspapers in England were small, and only fifteen million copies appeared during the entire year. Not many people read a newspaper, but a great many heard the news. Most people listened to the contents read aloud by someone else, partly because the stamp duty in Britain made newspapers expensive and partly because a majority of the audience was not adept with written English.

The proliferation of newspapers in the second half of the century brought a curious consequence. In a few decades sensation and scandal, traditionally oral, were newspaper matters. The spreading of rumors orally became thus associated with lack of rigor and only written sources were considered trustful. The importance of books as a way of advancing in a career can be also said to have originated in this century. Previously in history, literacy and learning were requirements to get a job in a fortunate field but it is not until the nineteenth century that manual jobs also favored a literate formation. Bailey (1996, p. 39) noted that

> Grinders or crammers […] formed schools to prepare young persons for the tests. The examinations often demanded a knowledge of theoretical matters where the jobs to be filled required only practical skills. Critics of the examination system noted that "it is perfectly possible to get a man who can spell without getting a philologist" (Cecil, 579), but philology was nonetheless part of the test, and, among other things, candidates for appointment were expected to know

> something about etymology, syntax and the history of English [...].

It is important to note in Bailey's quotation that we are dealing with the beginning of a trend that is still current. Today literacy is an indispensable requirement to get any kind of job, from the most to the least manual activity. Likewise we now take for granted that speakers should display not only a good command but also a reasonable degree of philological reflection about their native language. As it is clear from Bailey's words, these are only recent ideas originated less than two hundred years ago, even if they have become part of the core ideology in our society today.

The amelioration of artificial lighting was definitely a factor contributing to the increase of literacy during the century. The U.S. schooling system also encouraged literacy, as education was one of the main concerns of the state. This point is related to the emergence of English teaching, which also took place in the nineteenth century. Up to this century, English was not seen as a matter of scholarly interest. Grammars existed but they were rather aimed at the study of classical languages. English was not studied at universities. This situation changed progressively and not only human but also material efforts were directed to ameliorate the situation. As a consequence, libraries and buildings were built, publications were issued and teachers were hired (Bailey 1996: 12). This concern about language brought an exaltation of Standard English, which grew as a social expectation in the century. Bailey (1996, p. 13) noted:

> In the United States, for instance, literacy became a prerequisite in many states for full citizenship, whether for immigrants or for voter registrants. At the beginning of the nineteenth century, such requirements would have seemed to threaten the foundation of democracy and to deny the

> promises made to Europeans of many languages who were recruited to populate the United States.

The spread of democracy is also relevant in this field. Politicians' main weapon was, and still is, language. Standard language is therefore a requirement for democratic campaigning. It may be important to note, however, that literacy was seen useful only for freeborn male citizens, that is, only those who had the right to vote. Schooling and literacy for underprivileged social classes –black slaves in the United States or Irish and Cornish peasants in the United Kingdom– had an opposite connotation and teaching those people to read was actually regarded as a threat to society.

Another decisive technological change that affected language was the growth of communication. In this century there was an unprecedented growth of population mobility, which was possible thanks to the amelioration of machinery. Railways, steamships and motor vehicles replaced animal transportation and facilitated rural and urban communication. Concerning written communication, the amelioration of the means of communication meant a decrease in the price of postage and a maximization of its efficiency. People had it easier to send and receive letters. Similarly, the telegraph and, at the end of the century, the telephone, also opened new paths for communication. Bailey (1996, p. 59) commented on the effect that the telegraph had on language, as a new variety of English, more concise and elliptical, called 'telegraphese' originated. As is the case today with the use of elliptical English in text messages and in the Internet, telegraphese raised concerns about the purity of the language among conservatives. The way in which all these changes affected language is associated mainly with the potential speed in which language change could occur: an increase in communication could mean that a neologism, for instance, could reach more people in less time and become

standardized. All these processes, and many others, were taking place simultaneously and helped shape English as we know it today. However, it is perhaps the consolidation of English as a rightful language with a well-recognized standard variety the process that interests us more now as it is directly linked to the establishment of a scientific register. Even if early Modern English was considered, unlike Middle English, a language apt for commerce, politics, and general life, scientific communication in England before the seventeenth century had traditionally been carried out in Latin. The existence of a standard variety and the influence of the Scientific Revolution can be seen as the main facilitators for the creation of a scientific register in English.

1.3.1. Scientific language in the modern period

The birth of scientific English is a delicate matter. Most scholars refer to Halliday (2004), who signaled Newton's *Treatise on Opticks*, published in 1704 –and written twenty-five years before– as the first scientific text in English. His appointment, though pertinent, is dubious as he also cited Chaucer's *Treatise on the Astrolabe* (1391) as a very influential precursor. Similarly Taavitsainen and Pahta (2004) explored the continuist path and noted that English had already been the language of transmission of scientific knowledge since OE. As in the case of the Scientific Revolution, considered by Crombie (1974) and Hall (1954) a continuation of previous models, scientific register would have been simply readjusted, especially after the publication of the

Philosophical Transactions (Taavitsainen & Pahta, 1997; Camiña, 2013). According to Barber (1993, p. 214), the influence of science on language was to be seen not only in the expansion of vocabulary and the adoption of a particular lexicogrammar, but also in the use of a plain style written in prose as the most usual way of conveying not only scientific but also any kind of written knowledge: "The rise of scientific writing in English helped to establish a simple referential kind of prose as the central kind in Modern English". The establishment of English as the language of science was a pragmatic matter. Even if language, together with methodology and experimentation, was considered one of the pillars of the communication of scientific achievement, attitudes to language, as Camiña (2013) explained, were fairly negative at the time. Languages were considered defective, opposed to God's creations[18], and modern languages, a corruption from previous pure models. The decision to publish the *Philosophical Transactions* in English and not in Latin can therefore be regarded as a pragmatic attempt to reach a wider audience. Again, early members of the Society were compelled to work on their propagandistic skills to make room for their new understanding of science even if, in this case, language choice clearly contradicted their principles.

The plainness of language to which Barber (1993) referred was not a coincidence, as natural philosophers would deliberately search for a simplification of language that would "restore the biblical connection between words and things" (Camiña, 2013, p. 57). In their enterprise they also aimed at erasing the impediments to learning that had arisen as a result of the multiplicity of languages and the excess of ambiguity and

[18] As previously stated, the early period of the Royal Society was deeply religious, almost in a Puritan way.

redundancy[19]. In 1662, Wilkins' publication of the *Essay towards a Real Character, and a Philosophical Language* can be seen as an overt expression of the Society –or rather of Viscount Brouncker, second president of the Society, who commissioned the *Essay*– in its desire to achieve an educational reform based on the universalization of language performed by a purification of language. The *Essay* has been considered naive in terms of contents (Camiña, 2013, p. 96) but its influence on writers of science was great.

Movements and debates over language and the introduction of new vocabulary were common during the seventeenth century. Although the *inkhorn controversy* and the great debates on language had taken place in the fifteenth and sixteenth centuries, these movements contributed to set the mood for linguistic debate. The establishment of the Royal Society meant the beginning of the standardization of scientific language. In their study about pervasive communities[20] in the Royal Society, Allen, Qin and Lancaster (1994) showed how the enlargement of the scientific communities also meant their formalization. They (1994, p. 304) argued that whereas in the seventeenth and eighteenth centuries persuasive communities were relatively small and "scientific communication was relatively informal, with a great deal of personal communication" and the epistolary genre was indeed very common among scholars of that time, in the nineteenth century scientific communities began a process of internationalization and specialization.

On the other hand, the Royal Society served as well as a precursor to an academy of the English language. Even if in the end the idea of an institutionalized control of

[19] Even if the classical fallacy seems to have been abandoned, it is curious to realize that scholars still champion for a simpler, less ambiguous scientific language. In the case of nominalizations Billig (2008) is just one representative of the group of scholars advocating for a reduction because of the increase in the degree of ambiguity.

[20] Pervasive communities are made up by "scientists (and other authorities) whose works are cited in a particular corpus of scientific writing" (Allen, Qin & Lancaster, 1994, p. 279).

language did not succeed, the struggle over several decades to establish an academy of the language led to more self-consciousness over the language by scholars and general speakers. The eighteenth century were an age of nationalism all over Europe and America. National academies of science, such as the Royal Society in England or the Académie des sciences in France played an important role in the development of an organized and rationalized kind of technical science. Maier (2007, p. 19) claimed that "Enlightenment thinking and the constitution of nation states have put forward universalistic models of emancipation, such as human rights, liberty, equality and solidarity." In Maier's quotation it is easy to see reason for the establishment of a self-determined register for the sciences. If states claimed their sovereignty, so did science and it did it through institutionalization and development of a specific language register. This is the beginning of a dominant pattern of language register: the creation of an enterprise that portrays its power through specific language use. This could also be regarded as a quest for autonomy; science is not a state but it is also an entity and it calls for self determination following the tenets of the Enlightenment.

This chapter meant to represent a small account of the situation of English-speaking society, science and language. All the processes and events going on at the time caused the adoption and consolidation of nominalizations as scientific discourse markers. This historical dimension is key in my understanding of this linguistic feature. Today there is little doubt that English is the language of science in the world. As a result it must be used every day by millions of researchers who may or may not be native speakers. For the latter group, the use of nominalizations may not be as transparent as it is for native speakers. Many would even be required to read and write scientific texts in English even if they are not fully proficient in speaking and listening.

The pervading use of nominalizations as scientific discourse markers today is undeniable but in this chapter it has been shown that the situation was completely antagonistic 300 years ago. Their adoption was rather based on a shift of course in science. This shift was on the one hand an evolution in epistemological approaches to science but, on the other, it was also a revolution in the sense that new learned men sought for differentiation with regard to their antecessors. This conscious revolution resulted not only in the institutionalization and professionalization of the activity and in the banning of women from intellectual activities but also in the adoption of certain linguistic practices, among which we find this extensive use of nominalizations.

2. Nominalizations

This chapter is devoted to the description of the theoretical considerations regarding the study of nominalization. Although a great part of this chapter consists of a bibliographical revision of all the theories about nominalization provided by different linguistic schools, special emphasis has been put to create an original, synthetical approach. The approach here presented does not either drastically ally with or differ from any of the approaches described although it is also true that there are some points of discrepancy. The chapter starts with a reflection on the definition of nominalization and the problems associated (section 2.1). Very related to the definition lies the concept of transference (section 2.1.1), which for many linguistic schools is one of the defining features of nominalization. The study of nominalizations is here approached from two different angles. On the one hand, the description of the main theories concerning morphosyntax is contained in section 2.2. On the other hand, there is also an analysis of the functional implications of nominalizations (section 2.3). After both form and function have been analyzed, a special comment about the role of nominalizations as

discourse markers (section 2.4) is provided, before moving into the description of the typology created for this study (section 2.5).

2.1. Definition of nominalization

Nominalization is a well-known linguistic feature although scholars do not agree on their limits. Morphologically, there is discrepancy on the grammatical category of nominalizations because whereas the first studies on the matter focused on nominalizations as nouns, the tendency is to include not only lexical[21] but also clausal nominalization[22] (Comrie & Thompson, 1985; Downing, 1997; Heyvaert, 2003 ,2008, 2010; Koptjevskaja-Tamm, 1993, 2003; Mackenzie, 1985, 1986, 1987, 1996, 2007; Malchukov, 2004, 2006). Semantically, nominalizations were first believed to be transformations of verbs. Generativists did not provide a detailed account of the semantics of nominalizations. Later SFL scholars categorized nominalizations as the most common type of grammatical metaphor (Ravelli,1988) in which the meaning of process, which is normally encoded in VGs, is encoded in nouns (Albentosa, 1997; Banks, 2001, 2003, 2005a, 2005b, 2007, 2008; Guillén, 1998 Halliday, 1985; Ravelli, 1988; Ventola, 1996). Regarding the origin of nominalizations, some scholars refer

[21] The difference between lexical and clausal nominalizations is analyzed in section 2.2.1. However, broadly speaking, lexical nominalizations include nouns whereas clausal nominalizations are made up of those clauses, infinitival non-finite clauses and gerundial non-finite clauses.

[22] Malchukov (2004, p. 974) defined nominalization as "when a verb used in a NP function shows signs of decategorization and recategorization". Sušinskienė (2009, p. 84) claimed that "nominalization is a process by which a proposition (i.e. a congruent structure) is endowed with the function of a noun." Both definitions mark the nominal function as the defining feature of nominalizations. However, whereas Malchukov explicitly mentioned that clausal nominalizations were included in his study, Sušinskienė did not allude to this difference and from her definition, only nouns –that is, lexical nominalizations– are included. Nobody seems to exclude clausal structures from the definition of nominalization but actually unless the opposite is stated, studies focus on lexical nominalizations.

exclusively to verbs as the source of nominalizations[23] while others also focus on nominalizations derived from adjectives[24] (Sušinskienė, 2004, 2009, 2010a, 2010b, 2012). There are also scholars who claim that nominalizations do not derive from other categories, as they are independent ways of expressing a process or a state of things in a nominal form (Downing, 1997). Concerning the formation of nominalizations, most scholars tend to assume that there are three ways to form nominalizations. The first two –suffixation and conversion– are connected with the idea that nominalizations derive from verbs. Finally, there are cases in which nominalizations do not have an agnate verb (Banks, 2005b). This study focuses on lexical deverbal nominalizations that are formed by suffixation and indicate a process. It is assumed that all other options also belong to the configuration of nominalizations as a linguistic feature in English scientific register but they fall out of the scope of this study.

In this study, nominalization is understood as a linguistic expression of a conceptual representation of a process or state of affairs in a nominal form. This definition was built on the premises provided by Downing (1997, p. 147), who considers that situations and processes can be expressed linguistically in two major ways: finite clauses and nominalizations. Each of them has certain linguistic requirements and textual implications. Finite clauses are "likely to keep close to the speaker/hearer's experience of reality in terms of such features as chronological sequencing and agency" whereas nominalization allow "a presentation in terms of ideas, reasons and causality." This notion of nominalizations clearly takes a stand over the

[23] Within TGG (Chomsky, 1970; Grimshaw, 1990; Hazout, 1995; Jackendoff, 1975; Lees, 1960; Newmeyer, 1971; Siloni, 1997, Zucchi, 1993), nominalizations are understood as verbal transformations.

[24] As in the case of word category, nobody within the SFL tradition explicitly excludes adjectives as a possible source of nominalizations but unless stated otherwise studies focus on deverbal nominalizations, which are intuitively understood as the central category.

ubiquitous consideration of nominalizations as transformations. It is generally believed by most scholars that nominalizations imply some kind of transference, either morphological or semantic. Even if it is true that nominalizations create abstraction and are indicators of abstract thought, and some languages may have lower nominalization frequency rates (Koptjevskaja-Tamm, 1993), there is no reason to believe that verbs are closer to a supposedly more pure version of grammar in which nominalizations and abstraction are not prevalent. Indeed, one of the main hypotheses presented in this study is that nominalizations do not depend on verbs; they are different linguistic expressions of the mental configuration of a process. As such, they have their own structural and functional features. In scientific texts, they function as useful focalizers of information acting as linguistic guidelines for the reader to process information, which is granted by their valency flexibility potential.

The next pages are devoted to explain the concept of transference applied to nominalizations. Most theories revolve around it and therefore, the understanding of nominalization today is still deeply rooted in the idea that nominalizations are derivations from some kind of more pure, less abstract linguistic structure.

2.1.1. The concept of transference

Transference has been a key idea in the way scholars have understood nominalizations until very recently. The first insightful remarks on nominalizations in modern linguistics dealt with the idea that nominalizations are transformations of a verb into a noun

(Chomsky, 1970; Lees, 1960;) After an initial exclusive concern on the form of nominalizations, scholarly attention shifted to functions in texts in the 1980s. "Grammatical metaphor" was the expression used by functionalists (Halliday, 1985) to study nominalizations, as well as other linguistic features. The idea of transference, however, remained there, as all kinds of grammatical metaphors imply the transference of meaning of one unit into the form of another –in the case of nominalizations, the meaning of process usually codified in verbs is transferred into the shape of a noun. In the following pages, the concept of transference, as described by Transformational-generative grammar (TGG henceforth) and Systemic Functional Linguistics (SFL henceforth) will be briefly analyzed.

TGG theory is characterized by the assumption that a universal grammar categorizes language. The scope of study of generativism hence is not concerned with performance, but with competence and language is conceived to made up of two basic components (Chomsky, 1957, 1965, 1970). In the lexicon we can find the semantic information of the words existing in a language, as well as the phonetic and graphic representation of the realizations of those words. Grammar, on the other hand, deals with the relations between the words in the lexicon and it is composed of a base structure and a series of transformational rules that turn the base structure into other types of sentential or phrasal configurations. Nominalizations are indeed considered a transformation that turns a VP into a nominal form (Grefenstette & Teufel, 1998) and scholars have consistently tried to provide theories about it (Chomsky, 1970; Grimshaw, 1990; Hazout, 1995; Jackendoff ,1975; Lees, 1960; Newmeyer, 1971; Siloni, 1997; Zucchi, 1993).

In the nominalization process, the verb is ejected from its syntactic role into a nominal position. Thus, in (3)

(3) The enemy destroyed the city.

a full sentence made up of a verb (*destroyed*), its subject (*the enemy*) and its direct object (*the city*) is turned into a NP in (4)

(4) The enemy's destruction of the city.

where through the addition of a suffix, the verb becomes a noun (*destruction*) and is flanked by its former valencies, now turned into modifiers (*the enemy's* and *of the city*). This nominal form can be fulfilled by either nominal form of the verb (a deverbal noun) as in (5)

(5) The destruction of the city surprised the enemy.

or a gerundive as in (6)

(6) The destroying of the city surprised the enemy.

Concerning motivations for the use of nominalizations, generativists usually point out stylistic reasons like avoiding repetitions and awkward uses of verbs (Chomsky, 1970).

According to Banks (2005a, p. 78), nominalization can be described as a "form of grammatical metaphor whereby a process, which could be encoded as a verb, is encoded non-congruently as a noun." Other authors have produced similar definitions within the

SFL tradition (Albentosa, 1997; Guillén, 1998; Halliday, 1985, 2004; Ravelli, 1988; Ventola, 1996). Several terms, namely *grammatical metaphor* and *non-congruent realization*, are keywords in the understanding of the definition suggested by Banks (2005a).

The first concept, *grammatical metaphor*, recalls on the basics of the functionalist approach. Guillén (1998, p. 368) defined grammatical metaphor as "the transference of the linguistic representation of the semantic components of a situation between different lexicogrammatical categories". It may be noted here that embedded in the definition we find again the word *transference*, which echoes the concept of transformation studied by generativists. Again, nominalizations are believed to originate after some form of modification from a pure, more basic structure of language. In this case the tension is not produced between different lexicographic realizations but rather situated between semantics and lexis. Ravelli's definition (1988, pp. 135-136) categorized grammatical metaphor as "an alternative lexicogrammatical realization of a choice in the semantics." From these definitions it can be inferred that there are some prototypical realizations of the semantic components in terms of lexical categories and that grammatical metaphors disrupt this prototypical configuration and assign new linguistic realizations to semantic components.

The study of congruency, that is determining whether the function of a word is being congruently realized by its form, is one of the principles of SFL. Halliday (1985) acknowledged that there is a tendency to establish a one-to-one correspondence between the form and the function[25] of a word, which results in the endowment of meaning to the

[25] It may be noted here that in SFL *function* is associated with the semantic properties of a word, that is, with the semantical function that the linguistic unit has in the sentence. It must not be confused with the syntactical function of the word in the sentence.

form of a word and the establishment of a closed system of what he calls "congruent realizations of functions". These are typical, more natural ways of saying things and encoding functions into particular word class realizations. These more natural, congruent encodings are listed in table 1.

Function	Elements	Word class	Example
Process	(all types of) processes	Verb	*to make*
Participant	people, animals and objects	Noun	*maker*
Attribute	Events, mental and behavioral processes and objects	Adjective	*happy*
Circumstance	Time, place, manner, cause	Adverb, Prepositional Phrase	*always*
Relation	Relations	Conjunction	*while*

Table 1: *Congruent functional realizations.*

Clearly table 1 does not give a full account of all the possible matchings between word-class and function available in English. Those cases in which congruency is disrupted –for instance in a nominalization, where a process is encoded in a noun and not in a verb– are called non-congruent or metaphorical realizations. Despite non-congruent forms may be more frequent or even function as the norm, Halliday (1985) proclaimed the superiority of congruent realizations. Indeed, grammatical metaphors are widely used in adult speech and in some cases they have even become the default form.

Other SFL scholars have produced similar theories about congruency. Thus, in her article about the packing of information in scientific discourse, Ventola (1996)

established three basic ways in which function types can be encoded, namely the simple-congruent, the complex-congruent and the incongruent or metaphorical codings. The most basic way of packing information is using the simple, congruent coding, which relies on the congruent realizations shown in table 1. The complex, congruent coding increases the amount of information included in the sentence without altering its functional structure. The congruent nature of the unit is kept even though this does not imply simplicity in the understanding of the content. In nominal groups, this complexity is achieved by the addition of extensive pre- and postmodifiers or by means of rankshifted postmodifiers, whereas in the case of verbal groups, auxiliaries are used to increase its density. Finally, the metaphorical coding to which Ventola (1996, p. 182) referred is what "in systemic-functional literature [has] been called *grammatical metaphor*."

The relationship between the semantic component and grammatical categories concerning grammatical metaphor and the disruption of standard formulas was analyzed by Ravelli (1988). He set eighteen different categories of grammatical metaphor, which can be summarized in nine basic groups. These categories cover not only instances of nominalization –the first category in his classification– but practically all grammatical categories. Examples include cases like *for that reason* (type 5), in which a prepositional phrase functions as a logical connection –a function usually fulfilled by nominal or verbal groups–, or *historical experience* (type 9), in which an adjective replaces a PP and fulfills a circumstantial function. This classification also includes examples of rankshift (type 8) in which clauses function as phrases within a bigger clause, as in (7)

(7) [All it can do] is [to retaliate].

The existence of such a number of different kinds of grammatical metaphor does not minimize the importance of the process of nominalization itself, because nominalization is, indeed, "the type of metaphor of which there is the greatest awareness" (Ravelli, 1988, p. 140). The common feature for all the different kinds of nominalization can be said to be the convergence of the function of process and the grammatical form of noun. Thus, nominalization could be defined as a structure "in which verbal processes are coded in nominal structures" (Ventola, 1996, p. 153). Prototypically, the unmarked function of nouns is to express an entity or thing and that of verbs is to express a process. In nominalizations, however, the function of expressing a process is realized by a nominal group (Banks, 2005b). The process of nominalization could be systematized in the following way:

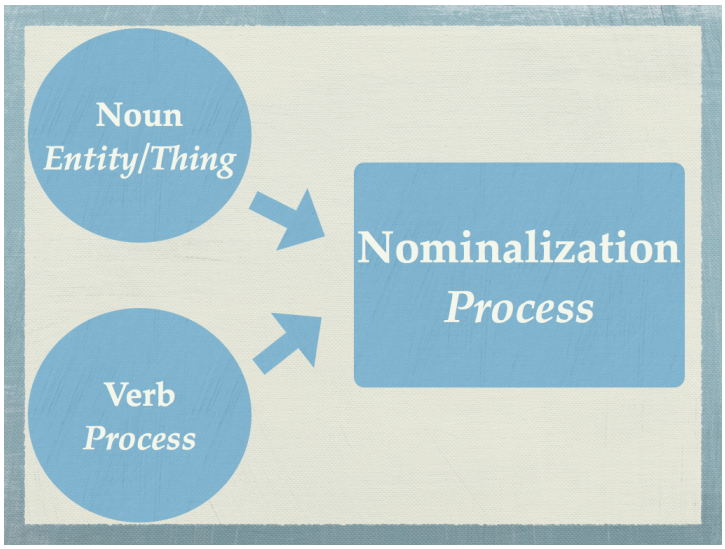


Figure 2: *Prototypical/unmarked functions of nominalizations.*

The merging of function and form results in one of the most characteristic features of this process. In nominalizations the semantic component of verbs, that of expressing a process, is encoded into a nominal group. In other words, nominalizations retain the semantic component of the *process* but they present it as an *entity*. Banks (2005b) compared grammatical metaphor to semantic metaphor. According to him, both grammatical and semantic metaphors are similar processes in the sense that both intend to achieve some effect on the discourse by the alteration of either the function or the meaning of a particular word. Thus, whereas semantic metaphor keeps the form of the word but alters its meaning, in grammatical metaphor the meaning is retained –with subtle modifications– but the form is changed.

However neither TGG nor SFL of approaches to nominalization were complete. The first focused exclusively on internal structure whereas the latter ignored structure and only studied functional implications. Dik (1985) and Mackenzie (1985) were very innovative in trying to join both approaches. Thus they conjoined functional readings with the study of the structure of nominalizations. Mackenzie (1985, p. 32) remarked that the study of nominalizations was centered on the relationship of the internal structure of the NP and its sentential equivalents, which in his view did not reflect the real use and structure of most nominalizations. Consequently he centered his study on the process of valency reduction. According to him, there are three main reasons why a speaker may have reduced the number of valencies by choosing a nominalization instead of a sentence: a) the avoidance of redundancy with the purpose of observing Grice's Maxim of Quantity (1975); b) the achievement of syntactic versatility and c) the introduction of discourse referents. After Mackenzie's turning, many scholars produced new theories about nominalizations which also took into account the premises of the

Optimality Theory (Prince & Smolensky, 2004) Although differing in minor concerns, nominalizations were believed to be the result of a conflict between lexical and functional constraints. In transcategorization processes, two independent processes are involved: on the one hand, the decategorization process (Hopper & Thompson 1980) implies the loss of morphosyntactic properties associated with the original primary function of the word (Malchukov, 2006, p. 974). On the other hand, the recategorization process assigns to an item some of the extended properties of the categories to which it functionally belongs (Bhat, 1994). Concerning the application of this process to nominalization, Malchukov (2006, pp. 1001-1002) claimed that

> Morphosyntactic properties of nominalizations arise from the interaction of constraints some of which are functional, some structural. On the functional side transcategorial processes are constrained by hierarchy constraints on deverbalization and substantivization.

As a result of the addition of a suffix, nominalizations lose verbal valencies in the deverbalization process (Givón, 1995) and the degree of transitivity of the VG is reduced (Albentosa & Moya, 2000). Paired with this process, the substantivization process assigns new functional properties to nominalizations (Fowler, 1991). All the studies about valency reduction and substantivization together with Mackenzie's allusion to Grice's maxim of quantity, Mackenzie (1985) made way to the understanding of nominalizations as linguistic expressions *per se a*nd not as transformations of verb*s*. It must be pointed out that today many scholars do not agree with this vision but its acceptation is becoming a standard. In the following section,

attention will be directed from the relationship between nominalizations and VPs to the actual morphosyntax of this linguistic feature.

2.2. Theoretical considerations about the morphosyntax of nominalizations

One of the main difficulties when it came to outlining this study was the lack of unanimity on the main point of study concerning nominalizations. Thus, whereas TGG has provided massive bibliography on the morphosyntax of nominalizations and has not been concerned with the functions the motivations in the use of nominalizations, SFL scholars have focused exclusively on functions and have given very vague structural information about nominalizations and their typology. This section is concerned with the morphosyntax of nominalizations and, consequently, many of the TGG readings on nominalizations are here analyzed. They include the difference in morphology between gerundives and derived nominals as well as the differentiation between event and result nominals. This morphological study does not only include the study of the noun by itself but also of the relationship of the phrase it governs and the VP (section 2.2.2).

2.2.1. Theoretical considerations about the morphology of nominalizations

Chomsky (1970) attempted to categorize nominalizations according to their form and their semantic and syntactic properties. He found that there were two main types of nominalizations and called them gerundive, that is *-ing* nominalizations, and derived nominals, which are the rest of nominalizations, not formed by *-ing* suffixation. Thus, all the sentences in (8)

(8) a. John is eager to please.
b. John has refused the offer.
c. John criticized the book.

can be turned into both gerundives as in (9)

(9) a. John's being eager to please.
b. John's refusing the offer.
c. John's criticizing the book.

and derived nominals as in (10)

(10) a. John's eagerness to please.
b. John's refusal of the offer.
c. John's criticism of the book.

Chomsky (1970, p. 30) described also their differences according to their behavior within the clause. Thus, he claimed that there are three main differences between gerundive and derived nominals:

1. Unlike gerundive, deverbal nominals are not very productive. In this context, productivity is taken as the propriety of being systematically applied to a wide number of structures. This is explained by the fact that not all verbs have a deverbal noun, while they do have an inflected *–ing* form that can function as the head of an NP.
2. The relationship between the nominal and its preposition is, in the case of derived nominals, idiosyncratic, whereas in the case of gerundives the preposition used is always that of the root verb.
3. Deverbal nominals have the prototypical internal structure of NPs. Gerundives apply some structural restrictions on their phrase. As a consequence, determiners and adjectives are not always admitted, as can be seen in the examples (11) and (12) proposed by Siloni (1997, p. 5)

(11) a. John's criticism of the book impressed us.
b. John's sharp criticism of the book impressed us.
c. The criticism of the book impressed us.

(12) a. John's constructing sailing boats impressed us.
b. * John's rapid constructing sailing boats impressed us.
c. * The constructing sailing boats impressed us.

Quirk et al. (1985) also distinguished between two types of nominalizations: verbal and deverbal. Verbal nouns are formed by the addition of the *-ing* inflection. They represent a description of the action taking place and are usually followed by support verbs because they still retain much of the semantic value of the verb they come from. Deverbal nouns, however, describe an action that has already taken place. Their

relation with the root verb is more unpredictable, either from a semantic, syntactic or morphological point of view[26].

Concerning the origin of nominalizations, TGG provided two different hypotheses: on the one hand, the lexicalist approach tried to demonstrate that nominalizations are the result of a movement that turns a verb into a noun. Nominalizations are, therefore, real nouns and even if their origin is verbal there is no reason to consider them transformations of verbs. Indeed, derived nominals are considered nouns in deep structure, not deep-structure transformations. The differences between gerundive and deverbal nominalizations lead Chomsky (1970) to adopt a lexicalist position and consider them two different types of nominalizations.

On the other hand, the transformationalist approach claimed that nominalizations are the result of a transformational movement that turns verbs into nouns. Acknowledging this transformation it is therefore possible to maintain the structural simplicity of languages. In transformationalist grammar, all possible irregularities and diversions from a universal simple grammar common for all human languages are systematically enunciated in a series of extensions and movements, among which nominalization has been widely analyzed. After Lees (1960), Newmeyer (1971) refuted Chomky's alliance to the lexicalist approach signaling that there was no sufficient evidence to prove that derived nominals could not be derived by a transformationalist rule in the same way that gerundives were. In spite of Newmeyer's remarks, the tendency in the following years was to progressively ascribe to the lexicalist approach

[26] Indeed, Grefenstette and Teufel (1998) recognized that deverbal nouns tend to establish metonymic associations and become concretized over time, consequently losing their original semantic value. To solve this, they decided to concentrate on *true nominalizations*, that is, nominalizations that retain their syntactic parallelism with regard to the original verb so that they could develop a method for automatically identifying support verbs for verbal nominalizations.

because, even if it was not so *elegant*, it could better explain the semantic differences between the two kinds of nominals.

Taking the semantic domain as the starting point, Grimshaw (1990) distinguished between two types of nouns that can form nominalizations: "result nouns", which express the product of an event and are usually associated with deverbal nominals and "event nouns", which denote an event or name a process and in many cases are found in gerundives. As Siloni (1997, p. 5) put it: "event nominals are the output of a process of lexical nominalization and gerunds are the product of syntactic nominalization". The morphological form of the noun does not seem relevant as there are examples in which the suffix *–ing* does not refer to an action that is taking place, but to the result of that action. Hence, in (13) and (14)

(13) The destruction of the city surprised the enemy.
(14) The destroying of the city surprised the enemy.

the change of suffix clearly marks the difference between result (*destruction*) and action itself (*destroying*). However, in (15) and (16)

(15) The painting was sold.
(16) The painting of the house turned out to be a chaos.

the same word (*painting*) expresses the meaning of both action and result and has to be disambiguated with the help of the context. The distinction between event and result nominalizations has attracted the attention of numerous TGG scholars. As Peris[27] (2012, p. 22) remarked, the two main points of discrepancy for the distinction of event and

[27] In an attempt to organize and settle the debate, Peris (2012, p. 32) established up to twelve criteria to distinguish result and event nominals. These include information about the verb (1), modifiers (3, 4) and former verbal valencies (5, 6, 7, 8), among others. Her approach is interesting but it falls out of the scope of this study.

result nominals revolve around argument capacity and lexical representation of nominalizations. Regarding argument capacity Grimshaw (1990) showed that the relationship between the nominal and its preposition (Chomsky's (1970) second claim) is explained thanks to the opposition event/result nouns. Grimshaw's claims were backed up by Zubizarreta (1987) and Siloni (1997, p. 3), who explained:

> Event nouns obligatorily have an argument structure as part of their lexical representations; they assign specific θ-roles, just like verbs. The lexical representation of result nominals, which do not express an event, does not specify an argument structure; result nouns do not take real arguments, which bear specific θ-roles, but rather a kind of semantic participants that are more loosely associated with them.

Some scholars (Alexiadou, 2001; Mel'čuk, Arbatchewsky-Jumaire, Elnitsky & Iordanskaja, 1984; Picallo, 1999; Pustejovsky, 1995) disagreed with this view and claimed that both event and result nominals can indeed incorporate part of the verbal argument structure into their phrases.

Concerning lexical representation, scholars debated whether nominalizations should be considered one or two lexical entries. One possible way of representation is to consider that nouns and verbs should be entered in the lexicon as unmarked items regarding the categories [+/-N] and [+/-V] (Alonso, 2004; Mel'čuk et al., 1984; Pustejovski, 1995; Zucchi, 1993). It is under the subcategorization of each of the subentries that the specification N/V is made, together with a list of all the idiosyncrasies of that item. Then, redundancy rules are created to express regularities of formation between items that have no apparent bond between them. A graphic representation of *attend/attendance* as one single entry is presented in figure 3.

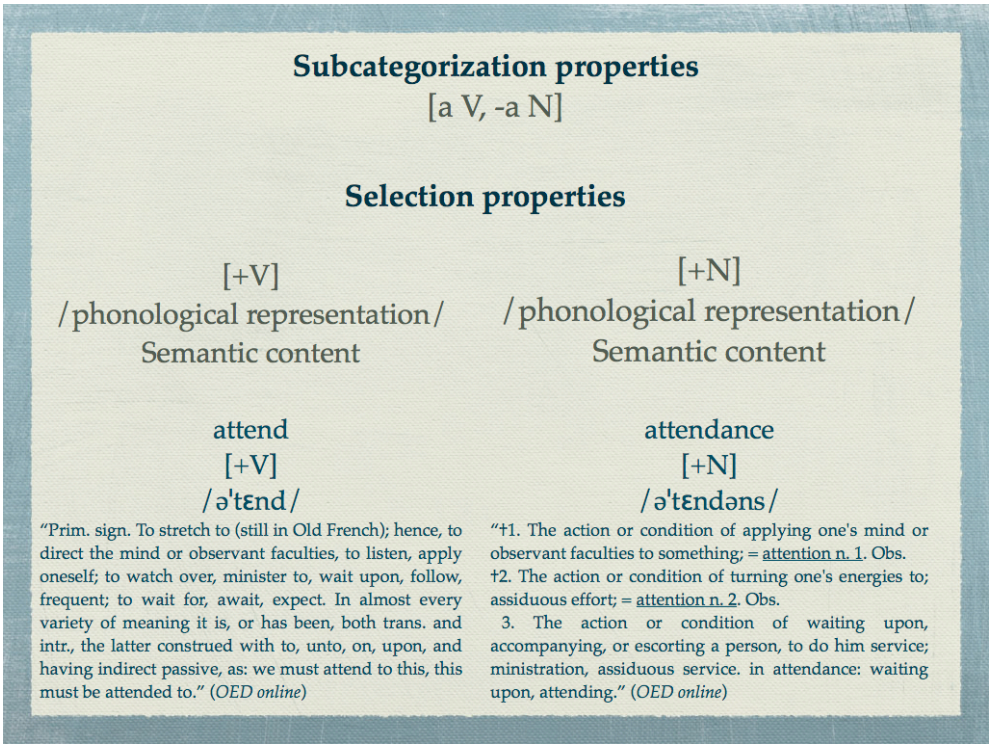


Figure 3: *Lexical representation of attend/attendance as one single entry.*

In figure 3 the entry is structured in two parts. In the first part, the description of the subcategorization properties [a V, -a N] indicate that this entry may be turned into either a verb or a noun. Then, in the second part, phonological and semantic information about all the possible lexical realizations. Figure 3 only includes the phonological representation and the definition of *attend* and *attendance,* but a more comprehensive representation would also have to include *attends, attending, attended* under the [+V] part and *attendances* and *attending* under the [+N] label.

The second option, proposed by Jackendoff (1975) and subscribed by Grimshaw (1990), Picallo (1999) and Alexiadou (2001) treated each of the items separately and included them under two different entries linked by redundancy rules. Each of these

entries completely describes the subcategorization of the items, as well as their phonological and semantic features, as exemplified in figure 4.

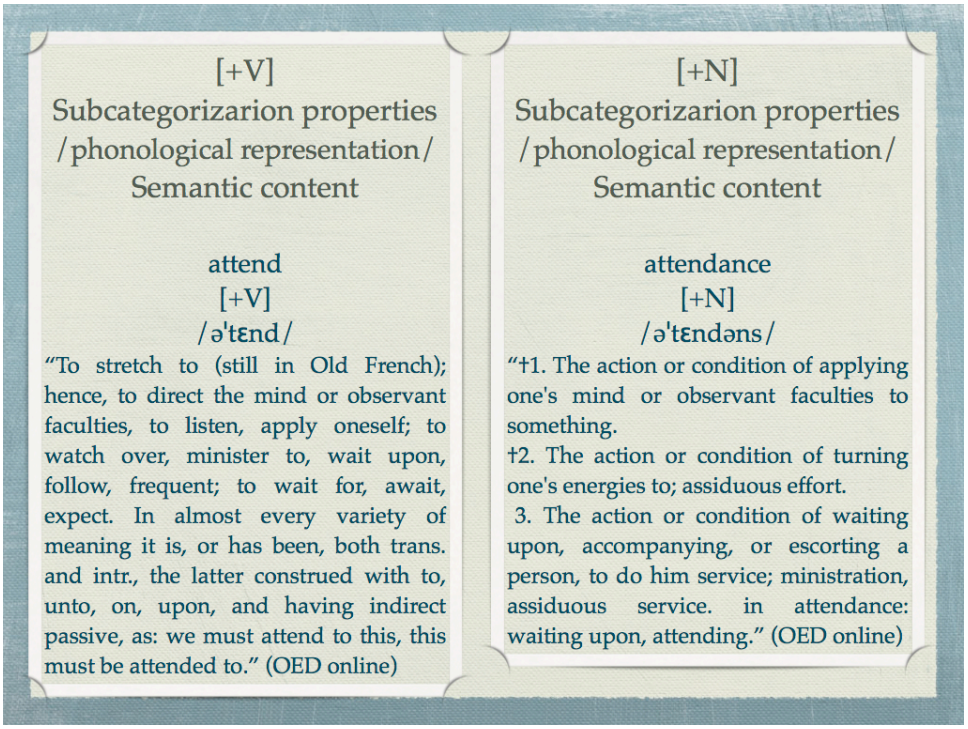


Figure 4: *Lexical representation of attend/attendance as two dictionary entries.*

One of the insourmontable points of disagreement with TGG is that in this study nominalizations are not believed to be simple verbal transformations. On the contrary, they are considered alternative ways of encoding a process. In this sense, in this study, nominalizations are considered to be closer to functionalist claims. Nevertheless, the difference between event and result nominals is loosely related to the typology proposed for this study and argument capacity is superficially addressed, as data analysis includes an analysis of the nominal modifiers in relation to the nominalization.

Concerning the morphology of nominalization, in order to comply with the definition of nominalization as a way of expressing a process in a nominal form, the first criterion is related to the actual form of the nominalization. Mackenzie (1985, p.

30) defined nominalization as any verbal predication that occupies a slot reserved for a nominal predicate. He based his definition on Lehman's (1988) study in which he treated subordination as nominalizations. Later studies have not been so inclusive but even if nouns are considered the central category, numerous scholars have widened up the notion of nominalizations to include also clausal configurations (Comrie & Thompson, 1985; Downing, 1997; Koptjevskaja-Tamm, 1993, 2003; Mackenzie, 1985, 1986, 1987, 1996, 2007; Malchukov, 2004, 2006;). Clausal nominalization can be divided in three groups. Firstly, that-clauses show very little difference with regard to the verbal configuration. In (17)

> (17) But obſerving the Sun, when the Earth comes to K, he will ſee the Sun with the Stars ♋, and will perceive **that he has chang'd his Place among the Stars**, and **that he has moved from ♈ to ♉** and ♊ to ♋, [...] (Gordon, 1726, p. 118; emphasis added).

even if it is true that both that-clauses could be replaced by a pronoun, the decategorization process is not noticeable. Aspect, voice, tense and mood, which are the first elements to disappear in a decategorization process (Malchukov, 2006), are still evident, and so are the participants, circumstances and agents of the *motion* process. This change is evident in non-finite clauses, like in (18)

> (18) [...] we ſhall here ſuppoſe the Sun **to move round the Earth both with a daily and yearly Motion**, as it appears to our Senſes (Watts, 1726, p. 3; emphasis added).

where the finite verb is replaced by an infinitive. Besides the reduction in verbal features, the agent (*the Sun*) is also alienated from the clause even if in this particular

case it is placed in an immediately preceding phrase, which simplifies its recognition. The same applies to gerundive non-finite clauses like (19)

> (19) But to the Sun, and all the other Planets (except the Moon) **moving round him**, he ſuppoſed carried about the Earth once in a Year, from [W]. to [E] (Morden, 1702, p. 6; emphasis added).

Lexical nominalization is influenced by decategorization processes but the effects of substantivization are more remarkable than in clausal typologies[28]. Even if lexical nominalization is, by definition, always fulfilled by a noun, this does not necessarily mean a reduction of valencies or the increasing of ambiguity. Thus, in (20)

> (20) The Ancient Aſtronomers obſerved of the Sun, that beſides **his apparent Motion round the Earth in 24 Hours**, by which he made, as they ſuppoſed, Day and Night (Harris, 1719, p. 26; emphasis added).

even the SVO pattern is kept, with the only exception of the circumstance (*apparent*) premodifying *motion.* Aspect, voice, tense and mood have been eliminated but it should not be difficult to reconstruct the action and its sequencing in time. Lexical nominalization can also show complete valency reduction and high levels of reification as in (21)

> (21) For action and its opposite re-action are equal, by Law 3, and therefore, by Law 2, they produce in **the motions** equal changes towards opposite parts (Luby, 1828, p. 11; emphasis added).

where little explicit specifications are offered about the agent and circumstances of the process. Ambiguity, however, does not seem to be an issue in this sentence because

[28] A deeper analysis of thematic nominalizations is offered in section 2.5.1.

given the topic of the text and the direct reference to Newton's laws of motion, it would be unmistakable for the intended audience that the agents of the motions are the celestial objects. Referential functions are thus maximized without compromising understanding.

The semantic component is also important, as they always express a process[29]. Other than processes, nominalizations can encode either the result of that process or one of the participants (instrument or agent) involved in it. However, only deverbal nouns that encode the process itself can be considered true nominalizations (Banks, 2005b, p. 348). Thus, in (22)

> (22) This abſurd ſyſtem puts us in mind of a **paſſage** in the 8th book of Milton's Paradiſe Loſt, where, ſpeaking of the ridiculous dreams of viſionary philoſophers, concerning the nature and motion of the heavenly bodies, he ſays, [...] (Bonnycastle, 1786, p. 58; emphasis added).

paſſage is more than a result of the verb to pass, as it has lost all possible semantic associations with any type of process and has become a wholly reified entity, which excludes it from a configuration as a nominalization. The same word in (23), however,

> (23) But on the shores of the larger continents, and where there are shallows and obstructions to the motion of the water, the interval between the time of the moon's **passage** of the meridian, and the time of high water, is very different at different places (Gummere, 1822, p. 237; emphasis added).

[29] The notion of process includes all events and most of result nominals.

still keeps the meaning of process and can indeed be considered a nominalization. Lexicographically, these are marked as two different subentries[30]. It may be noted that the structure may not always be, as will be discussed in the next section, a valid nominalization discriminator as in both (22) and (23) there is a PP modifying the nominalization but only in (23) the PP points out at one possible verbal argument (*of the meridian*). In (22), *in the 8th book of Milton's Paradiſe Loſt* refers rather to the location of the passage and has no relationship with the verb whatsoever.

The range of possible nominalizations is not restricted to nouns with a deverbal suffix. Conversion is another word-formation process that proves very useful in the construction of nominalized processes deriving from verbs. In (24)

> (24) HEVELIUS came after SCHEINER, and diligently watched the appearances of the ſpots for two years, the **reſult** of which application he has given us in this Selenographia and Cometographia (Wilson, 1773, p. 3; emphasis added).

reſult is a noun that has been formed using zero derivation. However, there is even no need in applying a word-formation process to derivate a noun from a verb, since there are also nouns that encode a process even if they do not have a cognate verb as in (25)

> (25) But to be ſerious, I ſuppoſe, Mercury and Venus being ſo near the Sun, have no **occaſion** to be lighted in the Night by Moons […] (Harris, 1719, p. 52; emphasis added).

As previously stated, only deverbal lexical nominalizations formed through suffixation have been considered because they are considered the central category in the

[30] According to the *OED online, passage* in (22) corresponds to the fourth subentry: "An episode or section", whereas *passage* in (23) can be contained under the fisrt subentry: "The action of passing, and related senses."

whole nominalization phenomenon. Apart of a review of the major contributions of TGG to the morphological study of nominalizations, this section has covered issues like the form, semantics and formation. The next section is concerned with the structure of NPs governed by nominalizations and, more particularly, with the relationship established between VPs and nominalizations.

2.2.2. Theoretical considerations about the morphology of the nominalization NP

Trying to find an explanation for the transformations and metaphorical deviations from a supposedly pure language structure, most scholars have systematically looked for similarities between verbs and nominalizations. Hence, if the mutation of a verb into a noun is normally performed through suffixation, all the elements of the sentence were thought to be able to be contained in the NP with a greater or lesser extent of morphological readjustment. Consequently, in (3) and (4)

(3) The enemy destroyed the city.
(4) The enemy's destruction of the city.

Chomsky (1970) proposed a minimalist portrayal of nominalizations in which the two verbal arguments are contained in the NP and the only marks of this transformation are the genitive Saxon and the preposition *of*. This offers only a slanted representation of nominalizations, as it does not really provide a realistic account of the structure of nominalizations (Mackenzie, 1985). Hopper & Thompson (1980) claimed that only 1%

of nominalizations retain two or more former verbal valencies and data extraction in this study have backed up their claim[31].

If instead of focusing on similarities, attention was drawn to differences, it could be easily seen that all nominalization modifiers are optional whereas verbal valencies are obligatory (Albentosa & Moya, 2000; van Dijk, 2008; Dik, 1985, 1997; Mackenzie, 1985, 1986, 1996, 2007). This fact is indeed of capital importance for one of the main claims in this study: that nominalizations have their own way of organizing information and that this organization is more flexible than in the case of verbal realizations and it also creates different effects on speech/writing. Behind all the functional implications this statement may entail[32], there is one basic feature of nominalizations as focalizers of information acting as linguistic guidelines for the listener/reader to process information. This is granted by their valency flexibility potential. Including the agent, the result, the object or a circumstance is not obligatory as in the case of verbal realizations, but rather a choice the speakers make according on how they understand the process and how they want to communicate it. Van Dijk (2008) trying to devilify the effect of agency deletion of nominalizations pointed out that nominalizations are not normally used in a conscious way but they are rather the consequence of a mental model speakers have projected into their minds on how they understood the process. Their high productivity is explained by pointing out that nominalizations are more lexically accessible

[31] See section 4.1.3.2 on valency transference.

[32] See section 2.3 for a detailed analysis of nominalization functions.

descriptions of an action that do not require valency inclusion and therefore, they are more economical[33].

Even if van Dijk's (2008) argument is sound, his explanation can account for agency deletion but does not really do it. It is mainly centered on the speaker and does not consider the receiver of the message. In this sense, Blakemore (1987, p. 95) claimed that the understanding of utterances is made from inferential computations and the use of certain focalizers of information may restrict those operations and these restrictions can lead the listener to expected effects and save efforts in processing[34]. Nominalizations are functional guidelines for information processing. They provide instructions for information processing and limit the potential ambiguity of the utterances. Much of this potential ambiguity is controlled in the inclusion of optional modifiers which give or omit information about the process. Thus, not only agency deletion, but also agency inclusion can be read as functional guidelines for correct information processing and this feature is exclusive of nominalizations[35].

Further evidence of the structural independence of nominalizations can be found in circumstance inclusion as well as in the role of APs in the NP. The inclusion of circumstances referring to process can be seen as an intentional way of focusing on

---

[33] He also pointed out other possible motivations for the use of nominalizations, namely lack of information or irrelevance about the agent, possibility of agent inference (either by common knowledge or from the text), lack of interest (due to shift of focus or to hide negative agency) and even lack of material space.

[34] This is related to the main tenets of Relevance Theory (Blakemore, 1987; Sperber & Wilson, 1995, 2006), which is based on Grice's theory. Sperber & Wilson (2006, p. 607) summarized the point of departure of the theory as follows:

> The goal of inferential pragmatics is to explain how the hearer infers the speaker's meaning on the basis of the evidence provided. The relevance-theoretic account is based on another of Grice's central claims: that utterances automatically create expectations which guide the hearer towards the speaker's meaning.

[35] See section 4.1.3.2 for an analysis of agency delition with corpus data and section 4.2.2.1 for an application of this variable to typological analysis.

particular points connected with the process. Unlike agents and results, circumstances are contained in the sentence as optional modifiers, whose influence over the process is to a greater or lesser degree inferior to that of obligatory valencies. This is reflected in the tendency to eliminate adverbs before verbal valencies as a result of the deverbalization process (Malchukov, 2006, p. 281). Similarly, in the debate about event and result nominalizations, scholars (Alexiadou, 2001; Grimshaw, 1990; Mel'čuk et al., 1984; Picallo, 1999; Pustejovsky, 1995; Zubizarreta, 1987) have usually cited argumentation capacity as one of the main discriminators, but they have disregarded adjuncts. Circumstance inclusion, however, may be considered a valid indicator of the independence of nominalizations regarding the VG. Since they lack the structural constraints of the VG, nominalizations may focalize attention on a circumstance rather than on the object or the result of the process through the inclusion of optional modifiers, which maximizes their functionality in information processing. In (26)

> (26) Every body perseveres in its state of rest, or of **uniform motion in a right line**, unless it is compelled to change that state by forces impressed thereon (Luby, 1828, p. 7; emphasis added).

the nominalization is only accompanied by two circumstances, one expressing manner (*uniform*) and the other, place (*in a right line*) and the proximity of the agent (*every body*) does not compromise ambiguity. At the same time, the distance of the agent creates detachment and facilitates objectivity[36] (Albentosa & Moya, 2000). Circumstances become the focal element, as they are the only element accompanying the nominalization. Restrictions concerning co-occurrence with former verbal valencies do not apply. Consequently in (27)

[36] See section 2.3 for a deeper analysis.

> (27) This decision was fully borne out by **[Dr]. Huggins's spectroscopic observation of the disappearance** behind the moon's limb of the small star Piscium, January 4, 1865 (Clerke, 1893, p. 324; emphasis added).

not only the agent (*Dr. Huggings*) but also the object of the observation (*the disappearance behind the moon's limb of the small star Piscium, January 4, 1865*) and the instruments implied in the process (spectroscope) are included.

The last piece of evidence to claim the independence of nominalizations concerning VGs is related to the role of APs in nominalization NPs. Concerning possible modifiers within nominalization NPs, some elements prototypically refer to the same verbal meanings. Thus, possessives and PPs introduced by *of* usually indicate the agent[37]. However, APs can fulfill multiple roles: they can indicate agents, as in (28)

> (28) For mankind muſt have made conſiderable advances in aſtronomical learning, before they could ſo far diſengage themſelves from the prejudices of ſenſe and **popular** opinion, as to believe in a doctrine ſo ſublime, and remote from **vulgar apprehenſion**, as that which the moderns have now firmly eſtabliſhed (Bonnycastle, 1786, p. 55; emphasis added).

where *vulgar*, as well as *popular* in the previous line, refer to people, the human agent that can perform the actions of apprehending and holding an opinion. APs can also refer circumstances about the process, as in (29)

> (29) While the general aspect of the planet reminded him of that of Mars. [...] but the difficulties in the way of its observation are enormously enhanced by its **constant close attendance** on the sun (Clerke, 1893, p. 304; emphasis added).

[37] See sections 4.1.2.1 and 4.1.3.1 for the analysis of possessive structures and valency inclusion in nominalizations found in the corpus.

where constant and close reference to the way in which Mars rotates around (attends) the Sun. Mackenzie (2007, p. 228) pointed out the apparent functional unpredictability concerning of APs and nominalizations although he specified that in double-possessive nominalizations, the role of adjectives is equivalent to that of adverbs, as they express a circumstance. In (30)

> (30) The next announcement of the discovery of "Vulcan" was on the occasion of the total solar eclipse of July 29, 1878. [...] This time it was stated to have been seen at some distance south-west of the obscured sun [...] and **its simultaneous detection by two observers** —the late Professor James [C]. Watson, stationed at Rawlins (Wyoming Territory), and Professor Lewis Swift at Denver (Colorado)— was at first readily admitted (Clerke, 1893, p. 307; emphasis added).

the double-possessive modification of detection is paired with the adjective *simultaneous* that refers a time circumstance around how the two observers detected the position of Vulcan during a total solar eclipse. Finally, APs may hold no relation with any verbal process whatsoever. In this case, they function as attributes to lexicalized nominalizations. Hence, in (31)

> (31) In truth, no natural theory has yet been advanced which will explain these lines, while **recent observations** furnish material that seems to render artificial construction probable (Lowell, 1895, p. 113; emphasis added).

the effects of relexicalization are quite evident if we take into account that *recent* does not express any temporal circumstance about the process itself but rather about *observations* as a reified process.

Generally, information selection in nominalizations responds less constrained principles in the case of nominalizations, either because of speakers' mental configuration of the process or because they want to control information decoding processes by focalizing or hindering information about them. In this section, three reasons –agency and circumstance inclusion and the role of APs– claiming the independence of nominalizations from VPs have been argued. Once the form of both nominalizations and nominalizations NPs has been studied, the next step is to consider their functional implications.

2.3. Theoretical considerations about the functional implications of nominalizations

The first remarks about the functions of nominalizations in speech were scarce. Chomsky (1970) claimed that they were mainly a stylistic choice to avoid repetition. Later, functionalists provided much more detailed attention to explain why verbal processes are encoded into a nominal form (Banks, 2003, 2005a, 2005b; Guillén, 1998; Halliday 1985, 2004; Ravelli, 1988; Ventola, 1996). Based mainly on the scientific and journalistic registers, most scholars highlighted the positive impact on texts, as nominalizations are believed to increase lexical cohesion and coherence of ideas, which eventually result in benefits for texts structure. Similarly, the substantivization process, which stresses abstraction and reification has been considered very positive for the

assimilation of new ideas. However, abstraction is usually correlated with valency reduction, and this has been signaled as an important ambiguity booster causing a great difficulty for readers. In the same light, CDA scholars have also interpreted nominalization as a vilifying linguistic structure that is used to hide agency, reproduce a certain ideology and perpetuate unbalanced power relationships (Billig, 2008, Fairclough, 1992; Fowler, 1991). Generally, the main functions attributed to nominalizations have been:

1. **Lexical cohesion (repetition, summarization)**: The nature of nominalizations is highly intertextual because they are typically formed using information that has been already developed. Thus, apart from functioning as a special kind of summary device, nominalized processes help develop a textual structure that contributes to the development of the main topic of the text. When constructing a text, not only is it important to find an appropriate internal structure to sentences, but a suitable balance has to be obtained so that the text can flow and its meaning can be easily decoded. Cohesive devices, such as lexical cohesion, reference, ellipsis and conjunction, play an important role here. In this sense, nominalization can be regarded as "a special kind of lexical cohesion" (Guillén, 1998, p. 370) because it is neither a repetition of what has been said nor a synonym and, at the same time, it is a combination of both. Marinkovich (2005) and Sušinskienė (2010a, 2010b, 2012) studied the role of nominalizations as lexico-grammatical cohesive devices and listed them as cohesive devices establishing sense relations in the text, defining cohesion and coherence and ultimately contributing to discourse organization. Based on Halliday and Hassan (1976) and Valeika (1985), she (2012, p. 132)

extended Guillén's (1988) notion of lexical cohesion by setting four types of cohesion: a) grammatical (reference, substitution and ellipsis); b) lexicogrammatical (articles, pronouns, conjunctives, conjunctive articles, particles, modal words, quantifiers and nominalizations); c) lexico-syntactic (periphrasis and parenthesis) and lexical (lexical repetition, synonyms, antonyms, superordinates, hyponyms, meronyms, paronyms and converses)

In (32) the alternation between verbal and nominal forms is evident and it is easy to observe the degree of lexical cohesion attained by the repetition of almost identical variations of the same lexical root.

> (32) When the moon is in the syzygies, her tendency to the earth is **diminished** by the attraction of the sun. The sun attracts both the earth and the moon, but with different forces, as one is nearer to him than the other; and as the attraction of the two bodies is mutual, whatever **diminishes** the attraction of the earth towards the moon, **diminishes** the tendency of the moon to the earth, and therefore in the conjunction and opposition of the sun and the moon, their tendencies will, by the action of the sun, be **diminished**. But the **diminution** in the syzygies is double to the **augmentation** in the quadratures. Had this **diminution** and **augmentation** been equal, at the octants the disturbing force of the sun would make no alteration in the moon's tendency to the earth, but as one is double to the other, the four points at which they balance each other, will be about 54°44' from the syzygies on each side.
>
> The **diminution** of the moon's gravitation to the earth in the syzygies, and its **augmentation** in the quadratures, tends to flatten her orbit in the syzygies, and to lengthen it in the quadratures; so that her orbit, even suppose it had been circular at first, is thereby made an oval or ellipsis, having the shortest diameter in the line of the syzygies, the longest, in the center. Hence the curvature of the moon's orbit will be greatest in quadrature, and least in syzygy (Ewing, 1809, p. 501; emphasis added).

The cohesive effect is clearly seen in the first paragraph, where *diminish/diminution* is repeated six times under three different forms. This way, the same idea is mentioned several times but the change in form does not make the text sound repetitive.

2. **Economy, conciseness and the packing of information**: nominalizations allow processes to be encoded in fewer words. This propriety may be useful in cases in which the writer wants economy to prevail over other considerations (van Dijk, 2008; Ventola, 1996, p. 183). Deixis and the avoidance of redundancy may be closely related to this point. Thus, when a writer wants to refer to a process and she considers that the specification of all the information related to it is not necessary, she may resort to a nominalization instead of a finite clause (Mackenzie, 1985, p. 33). This observes the Maxim of Quantity of Grice (1975) that states that to be informative it is necessary to include the necessary information, and no more.

Apart from being more economical, nominalizations present an alternative as far as lexical density and grammatical intricacy are concerned. Nominalized processes condense lexical words and increase the grammatical intricacy with regard to their verbal counterparts (Briones, Fortuny, Sastre & Botto de Porcovi, 2003). Thus, the coding of the information through a nominal or a verbal group depends on whether the writer decides to give preference either to grammatical intricacy or to lexical density to attain a desirable amount of conciseness in the text. Nominalizations allow the possibility that two processes are included in the same clause, one as the central process expressed by the VG and the other one being converted into a nominalized Participant, as Ravelli (1988, p. 145) explained:

> A ranking clause has only one process. […] However, if the process meaning is realized metaphorically as a Thing then it may function in the clause as a participant. In this way, two or more process meanings may be related within one clause, thus avoiding the clause complex systems. As a result, the ideational information of two or more clauses may be realized in one, with a correspondingly lower grammatical intricacy and higher lexical density.

Apart from pointing out the level of conciseness that a nominalization can confer to a text, Ravelli's claim serves to outline one of the main advantages of nominalized processes: the backgrounding of information.

3. **Backgrounding of information (theme, rheme and information structures)**: information structure is one of the main domains where the use of nominalizations makes a difference (Albentosa & Moya, 2000; Banks, 2005a, 2005b; Downing, 1997; Guillén, 1998; Halliday, 1985, 2004; Halliday & Martin, 1993; Ravelli, 1988; Sušinskienė, 2009; Ventola, 1996). As Halliday (as cited in Guillén, 1998, p. 371) claimed, nominalizations "made it possible on the one hand to construct hierarchies of technical terms, and on the other hand to develop an argument step by step, using complex passages 'packaged' in nominal forms as Themes." Thematization, a focal point development of information systems in the text, lies so close to nominalizations that Sušinskienė (2009, p. 87) claimed that they were a "morphologicotextual phenomenon" and Albentosa & Moya (2000, p. 461) referred to them as "architects of the written text."

The theme, the first element in the clause, states the point from which the clause is developed into the rheme. There is a narrow relationship between theme/rheme and

Given/New information. The basic function of these two information units is to organize the discourse in terms of information recoverable from previous parts of the text and new pieces of information. The function Given is assigned to what remains outside the range of the New. As Halliday (2004, p. 70) pointed out: "New is that which the speaker marks out for interpretation as non-derivable information, either cumulative to or contrastive with what has preceded; the Given is offered as recoverable anaphorically or situationally." In unmarked constructions the theme presents given information before proceeding into the rheme, where new information is included. In these constructions, themes are used as short summaries of what has been said before going on to give new information about the topic. Nominalizations are potentially very useful to meet this anaphorically recoverability quality that unmarked given themes seem to enjoy because they elaborate on previous contents. The intrinsic anaphoric property of nominalizations implies that what was a new rheme in the previous sentence becomes a given theme in the new sentence. Even if this violates the end-weight principle (shortest first, longest last) as Albentosa & Moya (2000) pointed out, scholars seem to focus rather on the fact that nominalizations are a "story in microcosm" (Bell, as cited in Downing, 1997, p. 147) and on textual implications.

The effects of nominalizations also tend to smooth the transition from clear-cut steps by merging them into a dynamic structure: the information of previous sentences is summarized into a nominal group and given as background information in order to facilitate the assimilation of new information in the second part of the clause. In (33)

> (33) The orbit of every planet is in a plane paſſing through the ſun, which planes are **inclined** to one another: thus in fig. 4. let ABCD repreſent the earth's orbit, or plane of the ecliptic; this is taken for a ſtandard, from which the **inclination** of each orbit of the planets, as EDFB, is meaſured. The **inclination** of the orbit of Mercury is 6°, 52' that of Venus 3°,33', of Mars 1°,52' of Jupiter 1°,30', and of Saturn 2°,30' (Adams, 1777, p. 5; emphasis added).

the backgrounding of information is evident looking at word class choice. In the first sentence, new information (that planes are inclined to one another) is introduced in the paragraph. Then, the process appears in a nominalized form (*inclination*) at the end of the sentence, which allows the assemblage of background and new information. Finally, in the next sentence, the same nominalization (*inclination*) appears again, this time as theme of the new sentence. The backgrounding of information is complete and now this is already given information that can develop into the rheme[38].

4. **Advancement of discourse**: this constitutes the logical consequence of the backgrounding of information. Once the rhetorical and informational structures of the sentences have been developed, it is reasonable that some functional meaning might be extracted from it. The methodic backgrounding of information through nominalized processes allows some degree of systematicity in the balance of backgrounded and foregrounded information. As Halliday (2004, p. 172) remarked, nominalizations are used "to create a discourse that moves forward by logical and coherent steps, each building on what has gone before." The dynamism of

[38] It may be noted, however, that in spite of its functionality, this use of nominalizations is mainly applicable to nominalizations in subject positions, because of their tendency to appear in the position of the theme. Even if scholars tended to focus on this type of nominalizations, data analysis has shown that subject is not the most common function for nominalizations. For more information see section 4.1.2.2, For the implications of syntactic function on typology see section 4.2.1.2.

nominalizations is not restricted to the limits of the clause in which they are used, since it also provides functional guidelines to facilitate the decoding of the text as a single unit. In this sense, nominalizations are valuable cohesive and coherence devices. In (32) nominalizations are used as lexical devices, summarizing the text and making discourse advance smoothly. In the first sentences of the paragraph, meanings of "process" appear in verbal encodings. Once the first nominalization is introduced in the fourth sentence, all processes appear in nominalizations and verbs become semantically empty. It is a different configuration of the sentence that responds mainly to stylistic and cognitive concerns. Cognitively, the second part of the paragraph is built on the foundations laid in the first part. If verbs are presented in a middle position in regular standard English, by putting them in a previous position we are copying the kind of movement we want to emulate in the reader's minds.

The reasons why writers might want to achieve this sense of advancement of discourse, which is related to their choice of nominalizations altogether, may be explained using some of the basic claims of cognitive linguistics. ICMs [39] provide stable ground for categorization of knowledge in our minds (Evans et al., 2007). Derived from sensorial experience, they help conceptualize reality and eventually they have some influence on language. Conceptual metaphors are a type of ICMs in

[39] Idealized Cognitive Models (ICMs) were first introduced by Lakoff (1987). They are similar to Fillmore's (1975) semantic frames, which are schematizations of experience, represented at the conceptual level that relate all the elements in a scene to human experience. Langacker (1987), in his theory of domains, also provided a similar definition and, like Lakoff (1987) and Fillmore (1987), he pointed out at the impossibility of separating lexical elements from their corresponding domains.

which a source domain is mapped onto a target domain[40]. This affects our conceptualization. Lakoff (1987) introduced a series of conceptual metaphors: LIFE IS A JOURNEY, CLASSICAL CATEGORIES ARE CONTAINERS, ACTIONS ARE TRANSFERS. The understanding of science in Western society is shaped according to a mix of all these, which could be summarized as SCIENCE IS A JOURNEY. Thus, the scientist is a traveler, purposes are destinations, methodologies are routes, difficulties are obstacles, counselors are guides, achievements are landmarks and choices are crossroads. As a result we do not understand science and knowledge as a static enterprise. On the contrary, we expect some kind of progression from them. In addition, time in English is conceptualized in terms of space (Lakoff, 2007, p. 280) in the conceptual metaphor TIME PASSING IS MOTION.

From this, it can be derived that time is things, the passing of time is motion, future times are in front of the observer and past times are behind him/her. This is connected to the advancement of discourse because, the longer we devote to reading a scientific text, the biggest progression we expect to have. Science develops, improves and turns what is new into the foundation of its subsequent progress. It never stops. It is this unceasing quest for constant progress and self-amelioration that catches the attention and imagination of scientists. This applies also to scientific works. Our expectations about scientific articles and books lie close to the idea of

[40] To see how ICMs affect our categorization of 'work' and how this is a language-dependant concept, see Bello and Müller (2011). In this study metaphoric and metonymic ICMs related to the expression of 'work' in idiomatic expressions in Spanish and German were studied. Data analysis showed that the body was the main source domain for the formation of cognitive metaphors because it allows to understand reality in terms of something already known. However, the study also showed that each language uses different values in work-related expressions. In Spanish responsibility, group membership, rivalry and flattery were the most repeated values whereas German favored organization and pressure.

"moving forward." Both learned and learning audiences expect some movement from "not knowing" to "knowing a bit more" when reading scientific writings. Hence the appropriateness of SCIENCE IS A JOURNEY. Given that ICMs pervade and shape our understanding of human experience, all the features associated with journeys are mapped onto our categorization of science as well and a series of correspondences are accordingly drawn.

In the case of scientific writing, the reader becomes the writer and the author of the book acts as the captain of the vehicle[41] that is transporting us to our destination. The tension created between verbs and nominalizations satisfies this expectation of forward progression. Visually, the advancement of the discourse is achieved with the use of verbs, occupying a middle position in the sentence combined with a use of nominalizations starting the sentences in the second part of the paragraph. The root of the word is kept, which clearly facilitates the task of assimilating both verbs and nominalizations as deriving from the same word. The movement from central to initial position has clear cognitive implications, as it accompanies readers in the process of assimilating new concepts and preparing them to build on those concepts to proceed further. Eventually the "moving forward" expectation is fulfilled.

[41] This mapping of education onto the domain of travelling/vehicles is, for instance, even more evident in the case of *coach*. According to the *OED online*, the first meaning of the word –that of "a large kind of carriage"– was introduced in English from Magyar *kocsi*, the genitive of Kocs, which was a Hungarian town specialized in this type of vehicles from the fifteenth century. The word was assimilated by a great deal of European languages. Thus, we have Spanish *coche*, Portuguese *coche*, Italian *cocchio*, German *kutsche*, Dutch *koets*, Bohemian *koč*. and Polish *kocz*. In English, however, the word underwent a process of semantic change in the nineteenth century, as it began to be used at the universities of Oxford and Cambridge to refer to private tutors that prepared students for examinations. This emphasis on personal assistance was then applied in the twentieth century to the world of sports, especially tennis, as coaches were expected to care about both the physic and psychological sides of professional players (Lureau, 2014).

6. **Abstraction**: Nominalizations are a result of objective thought. Unlike finite clauses, which are near the speaker/listener's perspective because they require chronological sequencing, tense and overt agency expression, nominalizations allow the presentation of abstract ideas and the expression of reason and causality (Downing, 1997, 2000; Eggins, 1994). According to Albentosa and Moya (2000), passives and nominalizations are clear objectivity facilitators causing abstract thought, which in the domain of child language acquisition is believed to be one of the markers of the transition between the last stage of cognitive development and adult thought (Inhelder & Piaget, 1958). In adult speech, this abstraction creates detachment, which has direct cognitive implications, as it allows to present processes as undeniable events.

Sušinskienė (2009, p. 84) claimed that

> Nominalization generally invokes an impression of abstraction in texts: they help us to create more distance between the even and the participants by removing the agents of actions.

She therefore linked the impression of abstraction to valency reduction. Her view is shared by many scholars that consider nominalizations one of the main transcategorial operations (Mackenzie, 1985; Malchukov, 2006; Sušinskienė, 2012) that consists of two independent, though complementary process: detransitivization (Albentosa & Moya, 2000; Givón, 1995; Hopper & Thompson, 1980) and substantivization (Dik, 1985). Detransitivization implies the loss of verbal valencies and other morphosyntactic properties of verbs (aspect, voice, tense and mood). Substantivization involves the acquisition of extended functions, that is the functions

of nouns, the category to which nominalizations properly belong. One of the effects of valency reduction is that it increases the level of implicit communication (Mackenzie, 2007) as a result of a different linguistic expression of a process which favors the formulation of reified ideas over inclusion of agency and chronological sequencing (Downing, 1997). The two following points are the consequence of the two processes involved in the transcategorial operation: detransitivization creates in most cases desagentivation and what has been called "mystification" by CDA scholars. On the other hand, substantivization results in reification, a useful feature for science transmission.

6. **Mystification**: nominalizations offer opportunities for deleting information about participants, time and modality to create certain ideological effects (Billig, 2008; van Dijk, 2006; Fairclough, 1992; Fowler, 1991; Wodak & Meyer, 2001). This is the main idea shared by CDA scholars, who side with hallidayan idea of congruency and grammatical metaphor. For them encoding a process into a nominalization is not random and it usually responds to ideology. After the process of syntactic reduction, nominalizations usually convey less information and turn whole sentences into agentless structures. When Lemke (1995, p. 60) claimed that "nominalization allows an entire activity, a process complete with its typical participants and circumstances, to be understood merely by naming it with the process nominalization", he was pointing out that this gap of information overtly conveyed in the text may be used to silence information about the agents and the circumstances in which the process took place and this almost always is done to maintain unequal power relations that are expressed in the ideology of texts.

Not all CDA scholars have shared the belief that nominalization use is always dictated by the intention of silencing agents and conveying ideologies that foster power relations and world inequalities. Van Dijk (2008) clearly criticized those who present nominalizations as a linguistic device used to reproduce unfair ideologies. He stated that the choice for a nominalization in a sentence is usually unconscious and it results from a process of combining the mental model that the speaker/writer has made about the process he/she wants to communicate and the context in which the communication is going to take place. There are a series of reasons why, after this mental processing of information, the speaker/writer may prefer a nominalization over a verbal realization, which can be summarized into two basic ones: economy and lack of information. Concerning economy, nominalizations need less obligatory complements and that makes them more lexically accessible descriptions of an action. This may be specially useful in writing where physical space may be an issue, for instance in headlines. There are other cases where the agent of the process expressed in the nominalization can be inferred –because it appeared before in discourse or in the text or because it is common knowledge. In these cases, the conciseness of nominalizations allows complying with the principle of economy and Grice's Maxim of Quantity (1975). Another reason why a speaker may consider a nominalization as a more effective linguistic choice is related to the lack of information about the agent or the circumstances in which the action took place. There may even be a lack of interest in the agent or its expression may be irrelevant. Here is where the mystification of nominalizations takes a stand. When a nominalization is used because there is a lack of interest in the agent, the focus is directed somewhere else. This may be to highlight the action conveyed in the

nominalization or to hide agency. A speaker/writer with specific intentions may then use this shift of focus and flexibility in valency expression offered by nominalizations to convey certain ideologies.

In a similar light, Martin (2008) also criticized Billig (2008) and extensively all CDA scholars emphasizing the mystificatory effects of nominalizations. For him (2008, p. 804) this is an opportunistic reading that "distorts the interpretation of the social function of grammatical metaphor in SFL research." Nominalizations are described as a resource for lexical extension that creates a tension between semantics –the expression of a process– and grammar –their incongruent codification as nouns. Similarly, he delimited the functions of nominalizations as markers of the acquisition of complex cognitive abilities as a result of sociolinguistic training. He (2008, p. 803) argumented that by secondary school, children are expected to manage the degree of abstraction nominalizations project to become literate members of society.

If mystification was one of the main consequences of agency reduction, the other transcategorial process, substantivization, is best explained in the reification process.

8. **Reification**: it refers to the attribution of factual, fixed properties to processes (Banks, 2005b, p. 349). When nominalized, processes tend to lose their dynamic nature and become somewhat more solid, static facts (Albentosa, 1997; Cadematori, Parodi & Venegas, 2006; Ciapuscio, 1992). In the process of substantivization, nominalizations acquire semantic features of nouns, the category they belong to. If the expression of a process through nominalization implies valency reduction (Mackenzie, 1985) and suppression of tense, aspect and chronological sequencing (Downing, 1997), "once verbs and adjectives have been nominalized they can be

talked about in more 'material' terms, as having taken place, occurred, etc." (Sušinskienė, 2009, p. 85), which Martin (2008) considers a lexical extension resource, indispensable for transmission of abstract thought. Nominalizations therefore are presented not as language specific but rather as a universal device related to the classifying of our experience of life that draws relations between things and processes (Sušinskienė, 2009). This can be put in relation to Fauconnier's (1994) mental spaces theory. Mental spaces are defined (Evans et al., 2007, p. 18) as "partial structures that proliferate when we think and talk, allowing a fine-grained partitioning of our discourse and knowledge structures." By reifying a process and presenting it as a material term, we are in some way creating a representation of reality and bringing it closer to our human experience. As reified entities, processes can be thought as agents or circumstances within other processes. This phenomenon can be seen as a disruption of congruency (Halliday 1985; Martin 2008) or a reflection of the dynamism of meaning construction processes involving the integration of structures across mental spaces (Fauconnier & Turner, 2002)[42]. In any case, what seems clear is that nominalizations are central to human thought and imagination.

The implications and effects of reification can be multiple. On the one hand, it may be effective at the persuasive dimension, as the formulation of processes as reified nominalizations favors their acceptance as immutable truth or, at least, possible

[42] The Conceptual Blending Theory (Fauconnier & Turner, 2002) is concerned with how we conceptualize knowledge and create meaning. Unlike SFL, infinite meaning construction is not seen as a disruption of a "pure" congruent form but rather as an integrative, dynamic process in which the emergent structure –in this case the nominalization used as participant or circumstance in another process– is more than the sum of its parts (Evans et al., 2007).

immutable truths (van Dijk, 1988; Downing, 1997, p. 152)[43]. This enables evaluation (Martin, 2008, p. 806), which can be either positive (Trew, 1979, as cited in Martin, 2008) or negative (Billig, 2008; Fowler, 1991). Once reified, "processes and qualities assume the status of things: impersonal, inanimate, capable of being amassed and counted like capital, paraded like possessions." (Fowler, 1991, p. 80), which offers opportunities for deleting information.

Another implication of reification involves the consideration of nominalizations as focalizers of information. In this study nominalizations are understood as functional linguistic guidelines for the listener/reader to process information. By drawing attention on the process, reified nominalizations are crucial for explaining relations between processes and "appropriately parcelling out information as peaks of thematic prominence, providing readers with an angle on the field and peaks of new building on from what can be assumed" (Martin, 2008, p. 804).

All the described functions can be applied in one or other way to all nominalizations. In some way, they all complement each other. Behind all these, the prevailing idea is that nominalizations are a concise way of expressing linguistically the conceptualization of a process. When compared with verbs, nominalizations can be more ambiguous due to valency reduction but they also provide valuable opportunities to organize discourse and express abstract relations among processes. They are an expression of adult abstract thought and, therefore, they are a very common linguistic feature in any type of adult speech. On top of that, nominalizations are now a well-

[43] It may be noted that van Dijk (1988) and Downing (1997) refer to journalistic register although it is believed that their claims can be applied to scientific register and, in some way, to any register.

known scientific discourse marker in English. In the following section, the adequacy of nominalizations to the register as well as some hints concerning their establishment as discourse markers are offered

2.4. Nominalizations as scientific discourse markers

The use of nominalizations as scientific register markers is not exclusive of the English language. However, the language of science has been a topic of major interest in the last decades and most works have studied the case of the English language or have based their hypotheses on it[44]. The difficulty of the language of science is usually paired with the abundance of specialized vocabulary at the lexical level and the adoption of certain lexicogrammatical features, namely passives and nominalizations (Halliday, 2004). Thus nominalizations have consistently been studied as markers of scientific discourse (Albentosa, 1997; Albentosa & Moya, 2000; Banks, 2001, 2003, 2005a, 2005b, 2007, 2008; Halliday, 1985; Halliday & Martin, 1993; Sušinskienė 2004, 2009, 2010a, 2010b, 2012; Vázquez, 2006).

According to Downing (1997, p. 151), nominalizations are found mainly on written genres because they tend to be abstractions that objectivize and stratify the process they refer to (Maynard, 1994). Downing (1997) linked the objectivization and

[44] For more information about the language of science, see Atkinson (1998), Beal (2012), Bello (2010a), Biber (1988), Camiña (2010, 2012, 2013), Camiña, Esteve and Lareo (2012); Crespo (2004b; 2012), Crespo and Moskowich (2005, 2009), Coffin (2006), Gallais-Hammonno (1981), Gómez (2013), Gopnik (1972), Gotti (1992, 2001, 2011, 2012), Gross (1990), Guillén (1998), Gutiérrez (2003), Halliday (2004), Halliday and Martin (1993), Lareo (2006, 2011), Lareo and Esteve (2008); Lareo and Montoya (2009), Lorenzano (2011), Martin and Veel (1998), Moskowich (2010a, 2010b), Moskowich and Crespo (2012), Sušinskienė (2004, 2009, 2010a, 2010b, 2012) and Swales (1990).

distancing created by nominalizations to Popper's (1974) "argumentative function." Even if Downing is referring to journalistic style, Albentosa (1997)[45] also referred to Popper in his analysis of nominalizations in scientific register. Drawing from Koptejvskaja's (1993) comparative study of nominalizations in seventy languages and Halliday and Martin's (1993) claim that the increasing use of grammatical metaphor is a manifestation of the augmentation of abstraction in modern scientific register, Albentosa (1997, p. 335) pointed out that nominalizations enable the evolution towards objectivity both in science and scientific register. The use of nominalizations meets the communicative needs of advanced societies. Progress towards abstraction and objectivity in science and society are compared with the child's cognitive development as described by Piaget and Popper.

Halliday (2004, p. 166) dated the birth of scientific English to the publication of Newton's *Treatise on Opticks* (1704) although by including an analysis of Chaucer's *Treatise on the Astrolabe* (1391), he was also making it obvious that the use of a nominalized discourse was also present before the establishment of a register as such. About the creation of a new register in English, Halliday (2004, p. 172) claimed that

> It is not that these grammatical resources were invented by scientific writers, What the scientists did was to take resources that already existed in English and bring them out of hiding for their own rhetorical purposes: to create a discourse that moves forward by logical and coherent steps, each building on what has gone before.

Scholars sometimes focus too closely on the adequacy of nominalizations for scientific –and also journalistic– register and seem to forget that, as Halliday noted,

[45] Also in Albentosa and Moya (2000).

nominalizations were already present in language well before the Scientific Revolution. Etymology[46] may also reveal that even if nominalization frequency rose in scientific register from the seventeenth century on, nominalizations were already a core element in the language. Indeed, Norman French was one of the main sources of nominalizations in English in the fourteenth[47] and in the sixteenth century the introduction of the printing press and the influence of the Renaissance also fostered the borrowing of more nominalizations –and words in general– from classical and Romance languages.It is therefore not very appropriate to link the abundant of nominalizations solely to an extralinguistic event like the Scientific Revolution.

On the other hand, abstraction is not an exclusive feature of modern science. The main criticism made to medieval scholasticism that triggered the Scientific Revolution was indeed the abusive use of abstraction. Influenced by Ockham's philosophy, Francis Bacon expressed the need for a reformation based on the separation between church and science, the adoption of an inductive method based on observation and experiment, and the use of a particular linguistic style avoiding abstraction[48]. It was the adoption of empiricism as a mainstream methodology that favored the increase in frequency of nominalizations as scientific markers. Thanks to valency reduction, nominalizations enabled writers to focus on experiments, not on agents, participants and circumstances. However pertinent with methodology nominalizations may be, their establishment as scientific discourse markers is rather related to the institutionalization of science in the

[46] See section 4.1.2.1 for data analysis following the etymological variable.

[47] The borrowing of words from Norman French started right after the Norman conquest (1066) but since nominalizations refer to abstractions, which are more loosely related to an actual physical conquest, their borrowing had a slow effect on language.

[48] See section 1.1.1.

eighteenth and nineteenth centuries[49]. The rejection of old scholastic approaches to science meant a shift of location from monasteries and universities to scientific academies, which in the eighteenth century went from a minority position (Voss, 1980) to virtually control the practice of science in Europe. The adoption of a particular linguistic style featuring specialized vocabulary as well as complex lexicogrammatic features like passive and nominalization could be therefore regarded as a group strategy to obtain and secure the new status of academics. Ventola (1996) considered grammar complexity a way of "guild codification", a code that only members of the community master and that distinguishes outsiders and novices from well-established members. By establishing a series of discourse markers, they could therefore control the access to information of outsiders distinctions and ultimately secure their position as professional scientists among peers.

Once established as a marker of scientific register, the tendency from the seventeenth century onwards has been to include as much information as possible into the foregrounded nominal group to facilitate the introduction of new information. Halliday (2004, p. 174) illustrated this tendency as illustrated in figure 5.

Halliday (2004, p. 155) asserted that the language of science in English has developed into more complex ways of nominalizing processes. The role of verbs is being progressively reduced because their function is no longer that of expressing processes or actions, but that of establishing a set of relations between the processes that are being expressed through nominal groups. Thus, not only nouns but also verbs are beginning to express functions that are not usually encoded by them.

[49] See section 1.1.2.

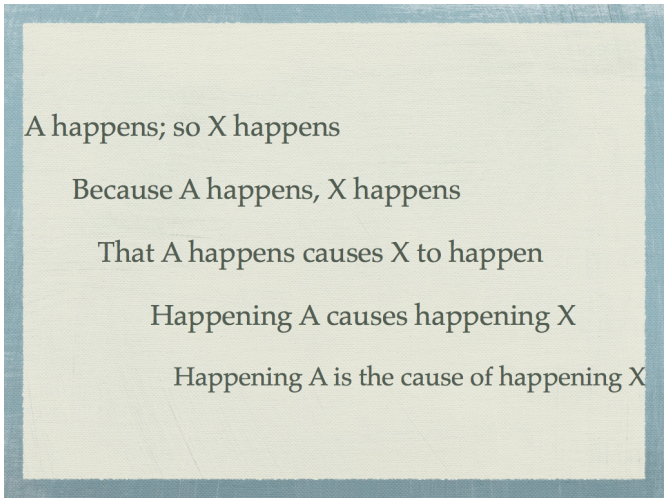


Figure 5: *Evolution of nominalizations in scientific English (Halliday, 2004: 174).*

This new role of verbs establishes a clausal pattern that has progressively become canonical –and therefore unmarked– in scientific register: each sentence contains two processes (one given, one new) expressed through nominal groups and linked through a VG. A schema of this pattern would look like:

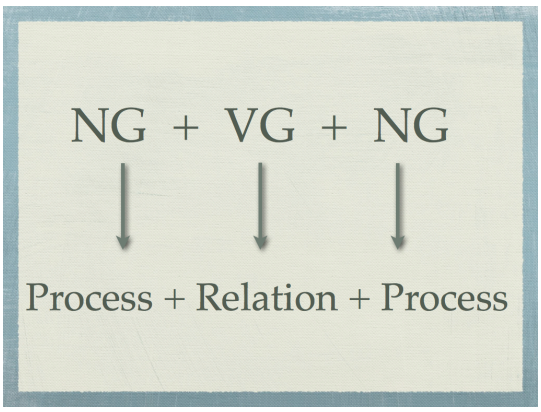


Figure 6: *Clausal pattern in English modern scientific register.*

Because of this, the current situation of the scientific language in English, Halliday (2004, p. 172) explained, is the result of a process that started 400 or 500 years ago, when the first scientific texts in English began to be published:

> Nominal elements in the clause are gradually taking over the whole of the semantic content, leaving the verb to express the relationship between these nominalized processes.

The tendency is to nominalize as much as possible and to increase of the relational aspect of nominalizations, or, as Halliday (2004, p. 174) claimed, there is a "steady drift towards the nominalizing region." But apart from its origins, the attributed factuality of reified nominalization proves very useful in the transmission and extension of scientific knowledge. because language becomes more compact, more direct to the specialist (Briones et al., 2003; Martin, 1993) The scientific activity is more and more conceptualized and comes to be perceived as an object (Banks, 2001, p. 11), which reinforces the unquestionability of the concepts exposed and facilitates their assimilation by the reader: "nominalization allows to present realities, facts or statements as unalterable or at least indisputable" claimed Albentosa and Moya (2000: 459; own translation) in response to Halliday & Martin (1993, 39, as cited in Albentosa and Moya, 2000, p. 459), who claimed that "you can argue with a clause but you can't argue with a nominal group." In scientific discourse, nominalizations are therefore a "resource for the construction and transmission of knowledge" (Halliday, 2004, p. 170).

Along with the functionality of reification, scholars have also highlighted the positive impact of the use of nominalizations on text structure and discourse organization. In this sense Martin (2008: 806) claimed that

> The language of science today is highly metaphorised, and conciseness of text has become an expected standard. We have learnt to encode dynamic processes as static participants and to pack a considerable amount of information into these static participants often in a very complex manner to a point in which nominalizations has become necessary to build knowledge and organize discourse.

Albentosa (1997) also referred to the impact of conciseness and Downing (1999) referred to nominalizations as "encapsulators of discourse", which is even more salient in the case of clausal nominalization. This adds to the consideration of nominalizations as cohesive devices contributing to coherence and discourse organization (Sušinskienė 2004, 2009, 2010a, 2010b, 2012). Again, the rhetorical purposes of nominalizations as concise structures that allow backgrounding of information and advancement of discourse are evident in Halliday's (2005, p. 169) words:

> Thus the device of nominalizing, far from being an arbitrary or ritualistic feature, is an essential resource for constructing scientific discourse. We see it emerging in the language of this period, when the foundations of an effective register for codifying, transmitting and extending the 'new learning' are rapidly being laid down.

In this quotation, apart from signaling the value of nominalizations for the packing of information and the construction of discourse, Halliday also referred to another important feature: that nominalizations were a linguistic feature chosen by a discourse community to fulfill a function. The establishment of the "new learning" also implied the establishment of a "new community of practice", who constituted a "new discourse community"[50] that adopted a "new language" (Gotti, 2013) About the establishment of the Royal Society, the main institution of the academicist movement in the U.K., as a community of practice and discourse, Gotti (2013, p. 282) claimed that

> [...] the group of natural philosophers that formed the Royal Society constituted a community of practice and discourse

[50] See Swales (1990) for a deeper understanding of "discourse community".

> who shared not only methodological aims and research activities but also linguistic conventions and discursive norms. In this way, they characterized themselves not merely as a sociolinguistic group but as a sociorhetorical one (Swales 1990:24), endowed with its own generic and stylistic features –all contributing to provide a common set of shared practices considered to be basic for the new members by this community.

In the same article he also claimed that, unlike other scientific or pseudoscientific societies like alchemists, members of the Royal Society emphasized the publicity of their work, which, of course, proved very important for the establishment of their linguistic practices as a new linguistic canon. Networking, another of the features of the Society, also secured that their practices became a standard. With this, it can be clearly seen to what degree the establishment of nominalizations as scientific discourse markers was not only a consequence of how functional nominal codings may be. Indeed, one of the most remarkable features analyzed in this section is that in spite of their functionality, nominalizations were chosen taking into account extralinguistic reasons and they were a result of the linguistic practices and academic aspirations of a new group of researchers. This is perhaps the most remarkable feature. Once the form and function of nominalizations and its role as scientific discourse markers has been analyzed, the typology of nominalizations used for this study is explained in the next section.

2.5. The typology of nominalizations

Again, as far as typologies of nominalizations are concerned, linguistic schools do not hold one single view. Generativists have mainly pointed out the difference between action and event nominals. Functionalists, on the other hand, have traditionally studied nominalizations in terms of the types of process involved in the semantics of the nominalization. Consequently, Ravelli (1988) developed a typology in which each of the six nominalizations was a non-congruent encoding of one of the processes described by Halliday (1985):

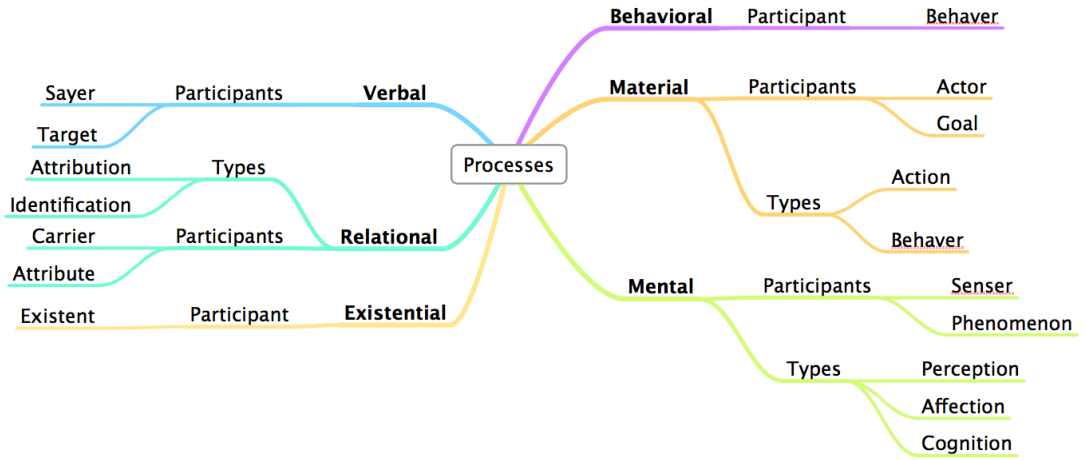


Table 3*: Process types and participants involved (adapted from Halliday, 1985, p 131).*

In this study, however, neither the pairing action/event nominal nor the typology based on process types was used. The main reason is that in both cases the notion of transference from a verb is evident. In this study nominalizations are defined as a linguistic expression of a conceptual representation of a process or state of affairs in a nominal form. It is true that verbs are also important in this notion of nominalizations

because they are another linguistic possibility for expressing the same (Downing, 1997). For this reason, the typology developed for this study was conceived as a continuum space in which processes may be linguistically expressed acquiring a certain degree of semantic and lexical features of either nouns or verbs, as can be seen in figure 7:

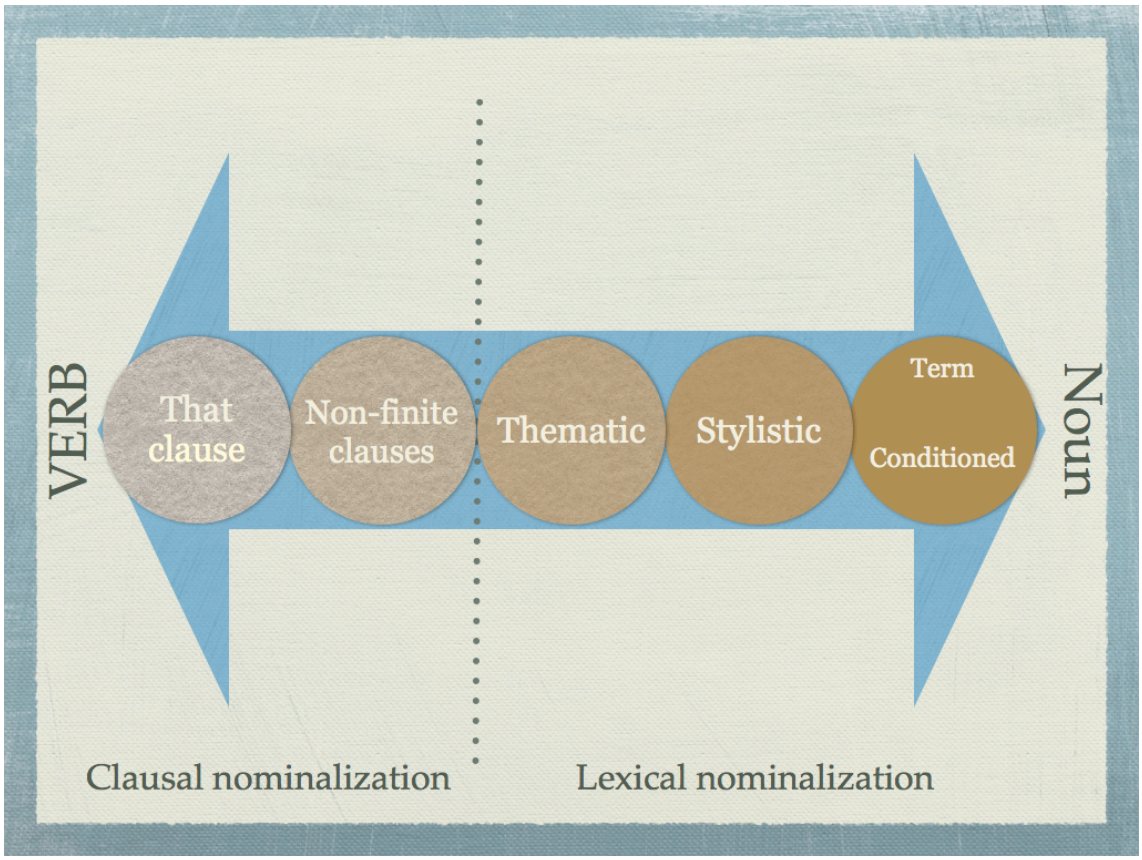


Figure 7: *Typology of clausal and lexical nominalizations.*

Depending on whether the nominalization shares more features with one of the poles, N/V, its situation in the scale will change. This typology is, in some aspects related to Malchukov (2006), who considers that nominalizations are a mismatch of two constraints in conflict: the lexis and the semantics of verbs and nouns. In the field of lexical nominalization, tense and mood do not apply but some morphosyntactic features

to be taken into account were the existence of modifiers, their size and their relationship with the agents and circumstances of the process. Similarly, the syntactic function of the nominalization was also included as a parameter and semantic considerations revolve around the degree of reification. Unlike Malchukov (2006) and Mackenzie (1985), textual considerations have also been applied. These include the rate of words from the same lexical field in the context –to determine to what degree the linguistic choice was free or context-motivated– and the role of the nominalization considering information structures. After the application of all these parameters, four lexical nominalizations have been distinguished in this study: thematic, stylistic, conditioned and term.

2.5.1. Thematic nominalization

This type of nominalization is usually found in contexts in which there is a high frequency of synonyms and words from the same lexical field in the same and preceding paragraphs, as in (34) and (35):

> (34) [...] then, the index remaining fixed, bring their limbs to the other wire, and if the ſame limbs be in contact, the axis is properly **adjuſted**; but if they lap over, the object end of the teleſcope is inclined from the plane of the quadrant, and muſt be altered by the **adjuſtment** for that purpoſe; but if the limbs be ſeperated, the object end is inclined to the quadrant, and muſt be **adjuſted** accordingly, and repeat the operation till the limbs coincide at both wires, and the **adjuſtment** is made (Vince, 1790, p. 15; emphasis added).

> (35) The time of high water is principally regulated by the position of the moon, and in general, in the open sea, is from two to three

> hours after that body has **passed** the meridian, either above or below the horizon. But on the shores of the larger continents, and where there are shallows and obstructions to the motion of the water, the interval between the time of the moon's **passage** of the meridian, and the time of high water, is very different at different places. The difference is so great, that at many places the time of high water seems to precede the moon's **passage**.
> For any given place, the time of high water is always nearly at the same distance from that of the moon's **passage** over the meridian (Gummere, 1822:, p. 237; emphasis added).

Indeed, both in (34) and (35) the nominalization appears only after the same process is expressed through a verb (*adjufted* and *passage*). In (35), once the nominalization (*passage*) appears, the verb is not seen again. However, in (34), the combination verb/nominalizations seems to aim at providing the text with some structure but avoiding lexical repetition. This structure relies heavily on thematic structures and the distribution of given and new information in the text[51]. Also in the nominalization the process is reified, which implies the acquisition of some of the semantics of nouns. Thus, its processing by the reader is simplified. Semantically, the proximity to verbal realizations is evident in this type of nominalization.

This is the type of nominalization Halliday (1985) had in mind when he developed his theory about nominalizations: a process that is codified into a VG (congruent codification) and works as rheme can be turned into a nominalization (grammatical metaphor) in the theme of the sentences at the end of the paragraph. Using nominalization as a device to pack and background information produces advancement of discourse. This nominalization also serves as a cohesive device, repeating and summarizing information. Sometimes it is merely a matter of avoiding lexical

[51] See again section 2.3 for more information.

repetition, as Chomsky (1970) remarked. If a verb appears multiple times, it is common that that verb may be replaced by a synonym verb or a nominal equivalent.

Structurally, this is the type of nominalization that has been most carefully studied. Thematic nominalizations are easily recognizable by their lengthy modifiers and their position in the sentence. They tend to include long pre- and postmodifiers that usually encode the agent, participants and circumstances of the process, as in the first nominalization in (22), *the moon's passage of the meridian,* where *passage* is flanked by the agent (*the moon*) and one of the participants (*the meridian*). Concerning premodification, thematic nominalizations have a higher tendency to be accompanied by demonstratives and possessives. In both cases, the function of the determiner is clear: in the case of the demonstrative, its deictic value usually refers to a verbal realization that appeared in a previous paragraph whereas possessives tend to encode the agent of the process. Concerning syntactic functions, thematic nominalizations are predominantly found in subject or direct object positions and usually paired with semantically-emptied verbs, which corresponds with the description of the canonical clausal pattern provided by Halliday (1985) for scientific register.

2.5.2. Stylistic nominalization

This type of nominalization is the result of the adoption of certain linguistic practices by a particular community of practice, that is, the astronomers of the scientific societies in the U.K. and the U.S. The choice of nominal over verbal realization in this type of

nominalization is made to meet stylistic concerns. Stylistic complexity is considered a way of guild codification (Ventola, 1996) that used nominalizations as complex markers of specialized discourse. In (37)

> (37) Her distance from the sun, like that of Ceres, is about 263 millions of miles, and she **performs her revolution** about it, nearly in the same time (Phillips, 1817, p. 68; emphasis added).

a simpler structure (*she revolves about it*) could have been chosen. However, the nominalization is inserted in a light-verb construction, which is, structurally, more complex than a simple finite clause. The choice of structural complexity in this case shapes the level of abstraction and specialization of the writer and helps delimit the intended audience. In this type of nominalization there is no textual reason involved. A verbal codification would be perfectly possible without adding complexity to the text. Also, the frequency of words from the same semantic family in the neighboring paragraphs is usually very low so the function of the nominalization as a synonym is not very plausible. This is obvious in (38):

> (38) Thus we have completed the proposed account of each individual planet, with the exception of the Earth and Moon, which have been deferred for reasons already given. If we were to pause upon the interesting and sublime facts which even this superficial view discloses, we should be filled with an **amazement** at their grandeur and beauty, to which no words can do justice, because we cannot either convey or entertain any adequate conception of their magnificence. In the planetary system, all is laid open to us, vast in design, and harmonious in revolution (Philips, 1817, p. 80; emphasis added).

The inclusion of the whole context in which the nominalization *amazement* is inserted is motivated by the interest of showing that there is no word related to either the semantic field or the lexical family of the word. The author could have written down *we should be amazed at their grandeur and beauty* instead of resorting to the nominalization. In that case, the nuance of "being filled" would be lost unless another expression was introduced. However, for stylistic reasons he favored the expression of the nuance in a nominalization.

Possible functions of stylistic nominalizations are associated to the status of the text as a product written by specific individuals –scholars– with a specific receiver in mind –the scientific community–. On the one hand, the use of complex linguistic devices echoes the level of desired cognitive complexity and abstraction: if a writer is able to produce a stylistically complex text, the latent implication may be that cognitively he/she will be able to attain high standards of abstraction and complexity in his/her area of study. Even if this syllogism is false, many scholars prefer a grandiloquent, wordy use of language[52]. Clearly the main aim of scientific language is communicating new knowledge and favoring further advancement, but authors may have other purposes, such as, displaying their command of the language to keep or attain a good position within the scientific community.

Structurally, a great deal of stylistic nominalizations are presented in light-verb constructions[53], which are a combination of a verb and a noun that form a single unit

[52] Many authors have recommended –and many still do– the adoption of a less pompous style for scientific register that includes less nominalizations (see, among others, Billig, 2008).

[53] The most common light-verb constructions found in this study were related to verbs of movement: *be in opposition, come into collision, make a motion/movement/revolution, perform a motion/movement* but there are also a goof number of verbs of saying: *give/afford an explanation, say a conjecture*.

with no extra semantic value (Alonso Ramos, 2004, p. 17)[54]. Light verb constructions are only one of the possible ways of expressing a verbal predicate, which could otherwise be also expressed with a single verb or with a copulative verb and a subject complement. Apart from stylistic concerns related to the adoption of a set of linguistic practices by a community of practice, one of the main peculiarities of this typology has to do with cognition and reification. As Banks (2001, p. 11) remarked, when nominalized, a process is increasingly conceptualized and becomes perceived as an object. Another motivation for this construction may be dictated by register, as some of the process terms –like *revolution* in (37)– are the key points of analysis in the discipline[55]. Thus, the light-verb construction would serve as some kind of textual cohesive device highlighting and summarizing important concepts to ease the assimilation of the text by readers.

2.5.3. Conditioned nominalization

Conciseness is one of the defining features of this nominalization. Usually forced by the rhetorical needs of the text, writers may favor to condense several processes into one

[54] The definition of *collocation*, or *light-verb construction*, has been the center of great scholarly debate (Alonso Ramos, 1994; Lareo 2006, 2008, 2009) which is avoided here because it falls outside the scope of interest of this study. It may be pointed out, however, that there are several trends in the understanding of what a collocation is. On the one hand, authors like Firth (1957) and Halliday (1966) based their definition on statistical criteria of co-occurrence in the same context. Then, lexicographers (Bally, 1951) introduced the notion of "lexical restriction" to, in some way, filter statistical criteria and separate collocations from mere usual combinations of words. Finally, scholars within the Meaning-Text Theory (Mel'čuk & Žolkovskij, 1970) highlight the importance of semantics and introduce the concept of "lexical functions", which amalgamate the semantic anf syntactic functions and valencies of both lexical units –the verb and the noun–. According to this school, in collocations, the nominal element has more semantic weight and selects a semantically-emptied verb.

[55] The main subject of study in astronomy in the eighteenth and nineteenth centuries dealt with the application of Newton's laws of motion to analyze and predict planetary and stellar motion and gravitation. It is no wonder that the nominalizations *motion, movement* and revolution may function in tests from this period as indexes of information.

single sentence. Conditioned nominalizations enable the omission of unnecessary information about participants and circumstances and maximize, while still retaining, the semantics of the process. In this sense, they may be less reified than term nominalizations, given that their main function is that of fitting into a particular context. In (39)

> (39) The same comet, also, came very near the earth; so that, had its quantity of matter been equal to that of the earth, it would, by its **attraction**, have caused the earth to revolve in an orbit so much larger than at present, as to have increased the length of the year two hours and forty-seven minutes [...] (Olmsted, 1841, p. 318; emphasis added).

the main objective of the author is to link two processes –that the comet attracts the Earth and that the Earth revolves in an orbit so much greater than the present– in a relationship of causality expressed by the verb (*have caused*). The agent and circumstances surrounding the first process can be retrieved from our knowledge of the world[56]. Thus, the decision to choose a nominalization over a finite process can be linked to a desire to condense and focalize the reader's attention on the process without compromising textual needs. Of all the functions of nominalizations provided by functionalists, only the packing of information can be considered intrinsic to this group. This kind of nominalization summarizes information to make it fit in a context in which more information is included. Besides, this type of nominalizations is used to simplify a convoluted grammar structure, providing a simpler structure and making the contents

[56] Indeed, they are explained by Newton's universal law of gravitation that states that any particle of matter in the universe attracts any other with a force varying directly as the product of the masses and inversely as the square of the distance between them." It is clear that this information was well-known for the intended audience of the text.

more easily accessible to the audience. Choosing a simple text structure may be motivated either by the semantics of the text –new or difficult concepts– or by the audience –learners and people not used to either the register or the concepts. The structure of this nominalization is highly unpredictable because it has to meet the requirements of the sentence in which it is included. Similarly, it is difficult to determine the reach of semantic shift in this kind of nominalizations because any nuance in meaning the grammatical metaphor may entail can be hindered by grammatical needs.

2.5.4. Term nominalization

The main distinguishing feature in this typology is related to the semantic codification of entities as nouns. The nominalizations in this group can be labeled as terms because they are totally reified. Terms are cognitive devices we create and use to study reality by establishing a set of differences and frontiers (Calvin, 1996; Eckardt, 1993; Lakoff & Johnson, 1980; Thagard, 1996). They are specially useful in scientific disciplines because they provide semantic traces of entity to both processes and entities themselves. The extensive use of terminology is, in fact, one of the defining features of the scientific register in any language. Thus, in (40)

> (40) But independent of theſe conſiderations, this rude ſyſtem was ſoon found incapable of ſtanding the teſt of **obſervation** and **experiment** (Bonnycastle 1786, p. 59; emphasis added).

the turning of a verb into a noun facilitates readers the identification of the processes and events that are being subject to study. Both *observation* and *experiment* are reified nominalizations functioning as guidelines to draw attention on the process. Information about agents is irrelevant here since the main objective is to present the processes as things. Cognitively, this process can be similar to the one employed when providing indexes at the end of a book to facilitate quick searches or when giving a title to a book or a chapter; It is related to a cognitive process linked not only to the reification of science but also to the organization of information in our minds.

Structurally, term nominalizations have usually underwent both a valency reduction (Mackenzie, 1985) and a substantivasubstantivizationtion process (Malchukov, 2006) so they are usually identified by the lack of semantic relationship between their modifiers and the agents, participants and circumstances in the process. According to Malchukov (2006, p. 976), these are clear examples of "strong nominalizations", that is nominalizations characterized by a lack of verbal properties and a total recategorization as nouns Concerning premodification, the most usual determiners are articles whereas postmodification is not common. Pluralization is also common in this typology. Syntactically, they can function in any position but they are the only typology found in titles given their extremely concise, reified nature.

This chapter has covered the most important features around the structure and function of nominalizations not only in scientific register but also in general language. At the morphosyntactical level, the only feature that all schools have highlighted is their ability to fulfill nominal positions. Their structure, origin and semantics have been continuously debated. From all the theories explained, I would highlight the fact that nominalizations have a particular way of expressing information about process.

Consequently, the inclusion of optional modifiers maximizes the function of nominalizations as focalizers of information and opens up a wide range of functional implications that include not only properties as discourse organizers but also as assimilation facilitators, which reinforces their value as tools for knowledge transmission. Concerning the typology presented in this study, it responds to structural and functional premises but it has also taken into account extralinguistic factors, such as the establishment of stylistic concerns motivated by new linguistic practices carried out by a new discourse community. Once established the theoretical framework used for this study, the next chapter will be concerned with the explanation of the corpus of study and the methodology.

3. Corpus and methodology

This chapter presents a description of the corpus of texts used for the analysis as well as the methodology used in it. Section 3.1 is concerned with the main work tools for this study, that is the corpus of texts and the search engine used. The description of the corpus is approached from two different angles: section 3.1.1 provides a general description of the corpus, explaining issues like size, textual categorization as well as sex, occupation and provenance of authors. Section 3.1.2 explores in detail the parts of the *CETA*, the subcorpus chosen for this study. Apart from general features of this subcorpus (section 3.1.2.1), an account of its different parts is provided, which also includes information about metadata files and prologues. The final part of this section is concerned with a brief explanation of the treatment of texts in the corpus. After corpus presentation, Section 3.1.3 deals with the description of the *Coruña Corpus Tool (CCT* henceforth*)*, the search engine used to retrieve information from the corpus in this study. Methodology is presented in section 3.2. Both the process of disambiguation and the

creation of the database used for analysis are explained. Additionally, I will expose the variables of study used together with the expected results.

3. 1. Work tools: *Coruña Corpus*, *CETA* and the *Coruña Corpus Tool*

Hickey (2003) reasoned the springing of diachronic corpora as a consequence of mixing together English historical linguistics with corpus linguistics, as a discipline in vogue after the initial hostility of generativists and the enormous advancement in computer science. Electronic corpora and the discipline of computational and corpus linguistics can be considered the turning point of linguistics in the last decades of the twentieth century, and its impact on the study of language can be paralleled to the impact of structuralism at the beginning of the century or the rising of generativism in the decade of the 1950s (Crystal, 1992, p. 85). Corpus linguistics deals with the principles and practice of using corpora in language study. The benefits of corpus linguistics revolve around a methodological reformulation that enabled to obtain quicker and more reliable data (Taavitsainen, 2005). This revolution in the method resulted in a shift of interest from random to central linguistic features, backed up by frequency numbers, which has led to the discipline of quantitative linguistics. The main criticism made to corpus linguistics was based upon the skewness of the discipline: "any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description [based upon it] would be no more than a mere

list" (Chomsky, 1965, p. 159). The debate over competence and performance continued for decades until the advancement of computer science made it possible to compile massive corpora, where Chomsky´s concerns about language underrepresentation became obsolete.

According to Tognini-Bonelli (2001, p. 2) "a corpus can be defined as a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis." There is not unanimity about the exact definition or specific features of a corpus. Some scholars allude to the etymological origin of the word *corpus* (Latin for 'body') for spreading the notion of corpora to any collection of texts. Hence, McEnery and Wilson (1996, p. 21) claimed that "in principle, any collection of more than one text can be called a corpus." Following this statement Kilgariff and Grefenstette (2003) studied the possibility of considering the World Wide Web as a corpus. This line of thought, shared by many (Ghani, Jones & Mladen'c, 2003; Robb, 2003; Sharoff, 2006), resulted in the creation of powerful automatic, web-based corpora, such as BootCaT[57] (Baroni & Bernardini, 2004) and Sketch Engine[58] (Kilgariff, Rychlý, Smrz & Tugwell, 2004). Although these corpora can be really useful, scholars generally agree that corpora should have explicit design criteria (Baker, 2002; Biber, 1993; Hickey, 2003; McEnery & Wilson, 1996 Oostdijk, 1991; Taavitsainen, 2005; Tognini-Bonelli, 2001). This could be taken as the main difference between *raw corpora* and *annotated corpora* (Taavitsainen, 2005, p. 326). McEnery and Wilson (1996, p. 21), who are usually taken as reference for the description of corpus design, highlighted four main specific features of corpora, namely sampling and

[57] http://bootcat.sslmit.unibo.it/?section=hom (Retrieved October 19, 2012).

[58] http://www.sketchengine.co.uk/ (Retrieved October 19, 2012).

representativeness, finite size, machine-readable form and standard reference, which will be analyzed in the next section.

The *Coruña Corpus, A Collection of Samples for the Historical Study of English Scientific Writing*[59] is compliant with all the specifications for corpus design stated by McEnery and Wilson's suggestions (1996). It is a closed corpus[60] with a finite size of around 400,000 words in each subcorpora and it is machine-accessible, thanks to a search tool (*CCT*) that has been designed for its joint use. The last prerequisite established by McEnery and Wilson (1996), standard reference, is also fulfilled, as the first of the subcorpora was released as a publication (Moskowich, Lareo, Camiña-Rioboo & Crespo, 2012), making it available to other researchers worldwide. Concerning Hickey's (2003, p. 4) suggestions about how to build a corpus, *CC* complies with all the requisites, as it is an untagged closed corpus presented in separate files. Additionally, normalization, one of Hickey´s concerns, has been carefully planned and different spellings were confronted with the *OED* and normalized, where applicable.

Still under compilation, the *CC* will be made up of several subcorpora containing samples of different disciplines according to the UNESCO classification of science and

[59] The *CC* is part of an ongoing project carried out by MuStE (research group for Multidimensional Corpus-based Studies in English) at the University of A Coruña. The main area of study of this group falls within the category of language variation and history of the English language and the common methodology for all members joins together traditional philological knowledge with new technologies. From 2003 to 2010 the group received funding to carry out the compilation of this project and, although the compilation of some of the sucorpora is still ongoing, the *CETA* subcorpus –the one used for this study– has already been published and others are ready for publication. In order to give coherence to the project and to widen the scope of the studies related to the corpus, the interests of the group have spread to cover other fields of knowledge. Thus, and as the result of the collaboration with the Information Retrieval Lab team at the Department of Computer Science at the University of A Coruña, a tool for retrieving information from the corpus (*CCT*) was designed. On the other hand, to understand better the scientific discourse produced by women, some of the members of the group are currently working on a project about women scientists from 1700 and 1930. The main aim of this project is to raise awareness about the contribution made by women to the field of science, not only as writers but also as assistants, editors, translators, illustrators and collectors, which were in many cases the only professions allowed to them (Crespo, Puente, Bello & Lojo 2012).

[60] McEnery & Wilson (1996) acknowledged that corpus can be either *open* or *closed*. Open corpora, also called monitor corpora are open entities to which updates are progressively being applied, whereas closed corpora have a finite size.

technology (1988). For this study I have selected one discipline from the field of exact and natural sciences –astronomy (*CETA*[61] subcorpus). All corpora in the *CC* have a common structure to facilitate contrastive studies about different subcorpora. In order to make the exploitation of the corpus easier, an information retrieval tool, called *Coruña Corpus Tool* (*CCT*) was developed. This tool has been especially designed for the *CC* by the IRLab (Information Retrieval Laboratory) at the University of A Coruña. The *CCT* enables the extraction of information of either morphemes, words or sets of words from the texts, which facilitates the study greatly[62].

3.1.1. Presenting Coruña Corpus

The *CC* has been designed to contribute to the diachronic study of English at several linguistic levels. The main idea behind this corpus is to study of scientific register in English. The project aims at complementing other corpora about to the history of English for specific purposes, such as the well-known *Corpus of Early English Correspondence*[63] (Nevalainen et al., 1998), the *Corpus of Early English Medical*

[61] See section 3.1.2 for a more complete description of *CETA*.

[62] See section 3.1.3 for a description of *CCT*.

[63] For further information and access to the corpus visit http://www.helsinki.fi/varieng/domains/CEEC.html (Retrieved November 20, 2012)

Writing[64], the *Lampeter Corpus of Early Modern English Tracts*[65] (Schmied, 1994), the *Heksinki Corpus of English Texts* (Kytö, 2012; Rissanen et al., 1991), and the *ARCHER* (*A Representative Corpus of Historical English Register*) (Biber et al., 1994).

Text selection was made according to the UNESCO classification of science and technology (1988). Marked with bold type in the following figure are the disciplines chosen for the *CC*, which so far belong to the groups of natural sciences, social sciences and humanities:

I. Natural Sciences.

Astronomy, bacteriology, biochemistry, biology, botanics, chemistry, entomology, geology, geophysics, mathematics, meteorology, mineralogy, computing, physical geography, physics, zoology and other allied subjects.

II. Engineering and Technology.

Engineering sciences such as: chemistry, civil, electrical and mechanical engineering and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; architecture, the science and technology of food production; specialized technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other allied subjects

III. Medical Sciences.

Agronomy, zootechnics, fisheries, forestry, horticulture, veterinary medicine and other allied subjects).

IV. Agricultural Sciences.

Agronomy, zootechnics, fisheries, forestry, horticulture, veterinary medicine and other allied subjects).

V. Social Sciences.

Anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), law, linguistics, management, political sciences, psychology, sociology, organization and methods, miscellaneous social sciences and interdisciplinary, methodological and historical S&T activities relating to subjects in this group.

VI. Humanities.

Arts (history of art and art criticism, excluding artistic "research"), ancient and modern languages and literatures, philosophy (including the history of science and technology), prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.), religion, other subjects and humanistic branches as well as other methodological and historical S&T activities relating to the subjects in this group.

Table 3: *Fields of science and technology (UNESCO, 1978).*

[64] The *Corpus of Early English Medical Writing* is composed of several subcorpora chronologically arranged: *Middle English Medical Texts* (*MEMT*) (Taavitsainen, 2005), *Early Modern English Medical Texts* (*EMEMT*) (Taavitsainen 2010) and *Late Modern English Medical Texts* (*MEMT*) (in preparation). Further information is available on http://www.helsinki.fi/varieng/CoRD/corpora/CEEM/ (Retrieved November 20, 2012)

[65] Further information can be accessed on http://khnt.hit.uib.no/icame/manuals/LAMPETER/LAMPHOME.HTM (Retrieved November 20, 2012)

The decision of not compiling any text from the group of medical sciences was, according to the compilers of the corpus a deliberate choice so as not to overlap with the *Helsinki Corpus*, a similar diachronic corpus texts specialized in medicine. Obviously, the *UNESCO* classification of science and technology (1988) was aimed at modern science and not all the texts compiled for the *CC* can be ascribed to one single category without serious doubts. The idea of science, as its language, has evolved with the passing of time and what was considered properly scientific in the eighteenth century may seem awkward for us (also *viceversa*!). However, the need for a clear-cut division and the lack of a standardized classification of science from the period covered in the corpus resulted in the adoption of the UNESCO parameters, as stated by the compilers of the corpus (Moskowich, 2012, p. 38).

The time-span of the texts compiled in the *CC*, 1700-1900, was chosen according to extralinguistic considerations (Moskowich, 2012, p. 47). Hence, the starting point of the corpus, 1700, coincides with a revolution of old epistemological patterns (Taavitsainen & Pahta, 1997). By 1700 the Scientific Revolution could already be considered an established phenomenon. The Royal Society had been running for 40 years and Isaac Newton (1643-1727) was already a senior scientist. Additionally, the beginning of the eighteenth century coincided with the start of the Enlightenment. As discussed in chapter one, the decline of the influence of religion over society together with the shift of scientific interest from deduction to induction gave way to a new form of science that favored the emergence of a new type of language for its dissemination (Swales, 1990). Several discoveries that took place at the end of the nineteenth century, such as J.J. Thompson's discovery of the electron, Planck's announcement of quantum mechanics in 1896 and Einstein's first formulation of the theory of relativity in 1905, set

a crisis in the basis of mechanical physics and serve as a good end-point for the corpus (Moskowich-Spiegel & Crespo, 2007: 348; Moskowich, 2012, p. 48).

The main aim of corpus design was to "construct smaller samples of the variety" (McEnery & Wilson, 1996, p. 21) to be studied, as analyzing every instance of language would be practically impossible. As a result, formal features include, apart from the external criteria for the delimitation of dates, equality in the sampling techniques and in the number of words per sample, and similar treatment of texts. *CC* contains two texts per decade and discipline. The general aim was to include one text from the beginning of each decade and one from the end. Samples contain around 10,000 words excluding figures, tables, formulae and graphs, which total ca. 200,000 words per century and discipline. Even if scholars in this field (Biber, 1993) have claimed that 1,000 word samples are long enough for the study of variation within the scientific register, compilers of the corpus (Moskowich, 2012, p. 39) have reasoned their choice of including larger samples as a result of historical reflection: the time-span of *CETA* covers a period in which the standardization of English scientific register was largely an ongoing process. Texts from this period present more variation and, hence, more words are needed to see repeating structures and patterns, as well as emerging standards of writing.

Issues of representativeness (Camiña, 2012, p. 96; Crespo & Moskowich, 2009; Lareo & Estévez, 2008, p. 70; Lareo & Montoya, 2008, p. 140; Moskowich-Spiegel & Crespo, 2007, p. 349) have been taken into account and include: use of first editions only and balance in text-types, gender and origin of authors. First editions of texts written in English by English-speaking authors have been preferred. Similarly no more than one text by the same writer has been included to avoid personal idiosyncrasies.

After text selection, the treatment of texts implied an edition of texts according to TEI standards, most widely used code for treating digital texts in the Humanities. Once edited, texts were saved in XML format. The corpus also contains other extralinguistic information about the authors in a series of XML files called metadata that can be consulted from the *CCT* interface. This information is especially useful to study extralinguistic variables. Regarding place and sex, the *CC* includes information about the place of education and sex of the authors. The audience, either real or intended, delimits the parameter of genre/text-type[66].

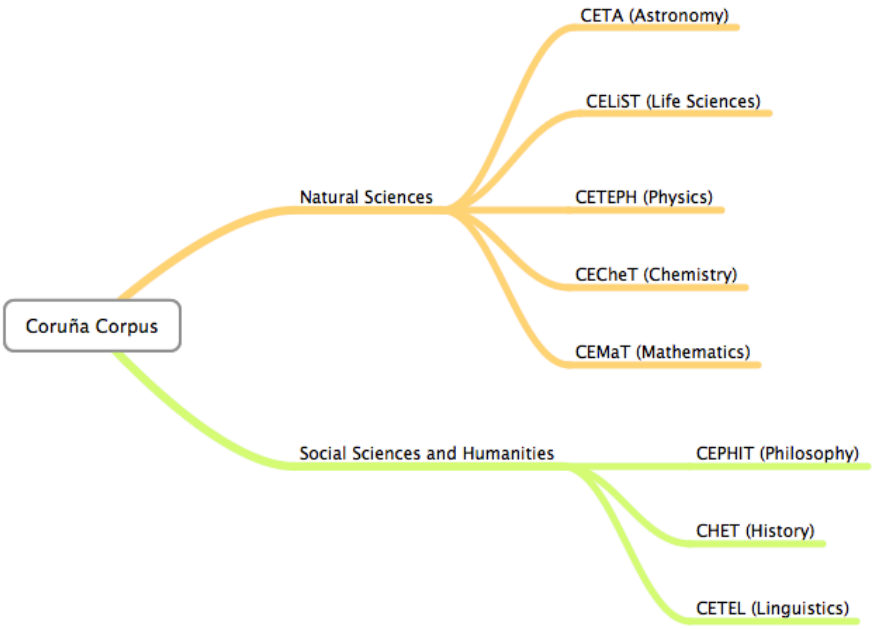


Figure 8: *Subcorpora in the Coruña Corpus.*

Several subcorpora were designed within the *CC*. Of all these subcorpora, only *CETA* has been published so far. *CEPHiT*, *CELiST* and *CHET* are ready for publication

[66] In this study, genre/text-type are understood in the way Taavitsainen (2001) did. She (2001: 140) defined *genre* as a "mental frame in people's minds which gets realized in texts for a certain purpose in a certain cultural context" and considered that text-types were a linguistic realizationbnof genres.

and the rest are still under compilation. The next section is concerned with a description of *CETA*.

3.1.2. *CETA*: *Corpus of English Texts on Astronomy*

As explained in the previous section, the *CC* will be made up of different subcorpora and *CETA*, the astronomy subcorpus, was the one selected for this study. Tables 4 and 5 give an account of the authors and texts used for this study. Additionally, these tables contain other data about texts and authors. Some of this information will be used as variables in the linguistic study, as will be explained in section 3.2. The information in tables 4 and 5 takes into account the information contained in the metadata files within *CETA* as well as in the descriptions provided by the compilers of the corpus (Moskowich & Crespo, 2012).

CETA authors had four possible nationalities: English, Irish, Scottish and American. Biographical research is an arduous work and it has been impossible to find information about all the authors in *CETA*. Following the information given by compilers (Moskowich, 2012, p. 49), six authors (Bradford, 1845; Charlton, 1735; Curson, 1702; Fuller, 1732; Gordon, 1726; Lacy, 1779) have been listed as "unknown" in the category "place of education." The reason behind this decision lies within the fact that no biographical information was found about them. For data analysis, however, these authors have been assigned a nationality taking into account other parameters than written records about the place of birth and education of authors. Consequently, Henry

Curson and Jasper Charlton were assigned English nationality based on the information about their careers: there is written evidence that Curson was a successful lawyer in the London area and Charlton is known to have worked as a customs officer at Gainsborough, Lincolnshire.

Period	Author	Date	Text Type	Place of education	Words
1700-1750	Morden, Robert	1702	Textbook	England	10 154
	Curson, Henry	1702	Textbook	Unknown	10 246
	Whiston, William	1715	Lecture	England	9 939
	Harris, John	1719	Dialogue	England	9 907
	Gordon, George	1726	Academic treatise	Unknown	10 437
	Watts, Isaac	1726	Textbook	England	10 407
	Fuller, Samuel	1732	Textbook	Unknown	10 232
	Charlton, Jasper	1735	Textbook	Unknown	10 358
	Long, Robert	1742	Textbook	England	10 474
	Hodgson, James	1749	Textbook	Unknown	11 106
1751-1800	Hill, John	1754	Dictionary	England	10 044
	Ferguson, James	1756	Academic treatise	Scotland	10 519
	Stewart, Matthew	1761	Essay	Scotland	12 180
	Costard, George	1767	Textbook	England	10 315
	Wilson, Alexander	1774	Research article	Scotland	4 240
	Adams, George	1777	Textbook	England	10 566
	Lacy, John	1779	Academic treatise	Unknown	5 908
	Nicholson, William	1782	Academic treatise	England	10 268
	Bonnycastle, John	1786	Letter	England	9 975
	Vince, Samuel	1790	Academic treatise	England	10 540
	Bryan, Margaret	1797	Textbook	England	10 263
TOTAL					208083

Table 4: *Authors and texts contained in CETA (eighteenth century).*

The nationality of the remaining writers was assigned taking into account the place of publication of their works, that is, England in the case of George Gordon, James Hodgson and John Lacy and Ireland for Samuel Fuller. I am aware of the possible inaccuracy of these assumptions but my belief is that this is the most plausible solution to make a full use of all corpus texts.

Period	Author	Date	Text Type	Place of education	Words
1700-1750	Small, Robert	1804	Academic treatise	U.S.	10 435
	Ewing, John	1809	Lecture	U.S.	9 985
	Brewster, David	1811	Academic treatise	Scotland	9 824
	Phillips, William	1817	Lecture	England	10 277
	Gummere, John	1822	Textbook	U.S.	10 507
	Luby, Thomas	1828	Academic treatise	Ireland	10 704
	Herschel, John	1833	Academic treatise	England	10 224
	Garland, Landon	1838	Research article	U.S.	9 608
	Olmsted, Denison	1841	Letter	U.S.	8 742
	Bradford, Duncan	1845	Textbook	U.S.	10 268
1751-1800	Barlett, William	1855	Textbook	U.S.	10 858
	Whewell, William	1858	Letter	England	10 079
	Mitchel, Ormsby	1860	Academic treatise	U.S.	10 183
	Loomis, Elias	1868	Textbook	U.S.	10 323
	Chauvenet, William	1871	Academic treatise	U.S.	9 895
	Steele, Joel	1874	Textbook	U.S.	9 979
	Young, Charles	1880	Research article	U.S.	5 181
	Darwin, George	1880	Research article	England	6 454
	Croll, James	1889	Academic treatise	Scotland	9 390
	Clerke, Agnes	1893	Academic treatise	Ireland	10 530
	Lowell, Percival	1895	Research article	U.S.	8 531
TOTAL					202403

Table 5: Authors and texts contained in CETA (nineteenth century).

The following pages will contain an in-depth analysis of *CETA*, starting with some of its general features, namely representativeness, size and textual categorization. The last part of section 3.1.2.2 includes information about the sex, occupation, origin and education of authors. The following section (3.1.2.1) is concerned with the different parts of the subcorpus, which include metadata files, prologues and sample texts contained in the subcorpus

1.2.1. General features of *CETA*

Representativeness and size are two of the more important parameters in corpus linguistics (Hickey, 2003; McEnery & Wilson, 1996;). Sample texts in *CETA* have an estimated size of 10,000 words. There are instances of texts with slightly smaller number of words as a consequence of text availability, as can be seen in figure 9. In these cases, another shorter sample text has been introduced in the same decade slot. Consequently, the decade of the 1770s contains three texts (Adams, 1777; Lacy, 1779 and Wilson, 1774), and so does the decade of the 1880´s (Croll, 1889, Darwin, 1880 and Young, 1880).

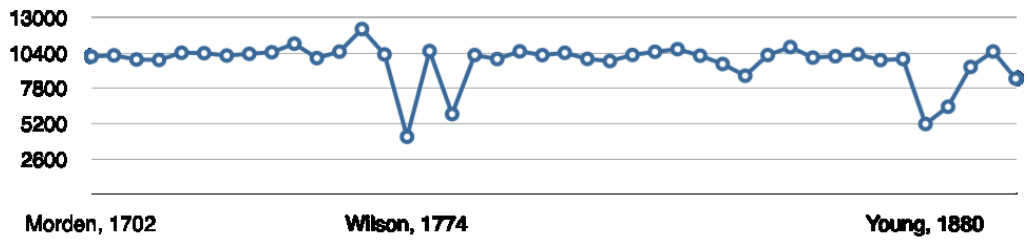


Figure 9: *Number of words in CETA according to chronology.*

Together, the total number of words remains balanced, with a total of 208,083 words for the eighteenth century and 202,403 for the nineteenth, as reflected in figure 10.

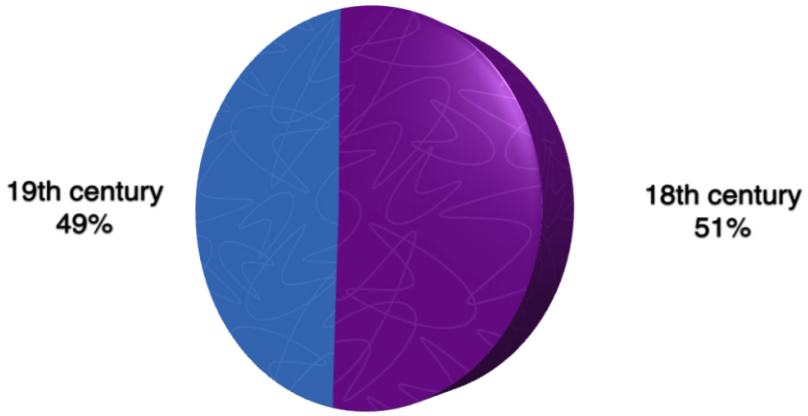


Figure 10: *Distribution of number of words according to century.*

According to the guidelines provided by corpus design theorists (Biber, 1993), *CC* is a representative corpus. "Sample size is not the most important consideration in selecting a representative sample," claimed Biber (1993, p. 243) and, he argued that a solid method of sampling aiming at representing "the full range of variability in a population" would be more important.

Textual categorization was another key factor for the building of the corpus. Samples in *CETA* have been divided into eight groups according to text-type, namely, essay, research article, letter, dictionary, academic treatise, dialogue, lecture and textbook. Compilers of the corpus based their text-type decisions in the notion of paragenre, that is, "a genre belonging to one professional community" (Moskowich, 2012, p 37). By doing so, they highlighted the sociolinguistic orientation of the corpus. *CETA* aims at compiling all forms of communication about astronomy among astronomers in the eighteenth and nineteenth centuries. Compilers clearly stated that their intention was that of collecting samples from different epistemological levels

covering from the high level of research articles to the apparent low epistemological level of textbooks.

Another very important issue in textual categorization regulating linguistic choices concerns the audience of the text. In academic treatises and essays, originally aimed at the scientific community, the expectation is to find a higher degree of complexity and nominalizations than in texts intended for a less learned audience, such as lectures and textbooks, among others. Figure 11 illustrates the distribution of text categories in the texts of the corpus.

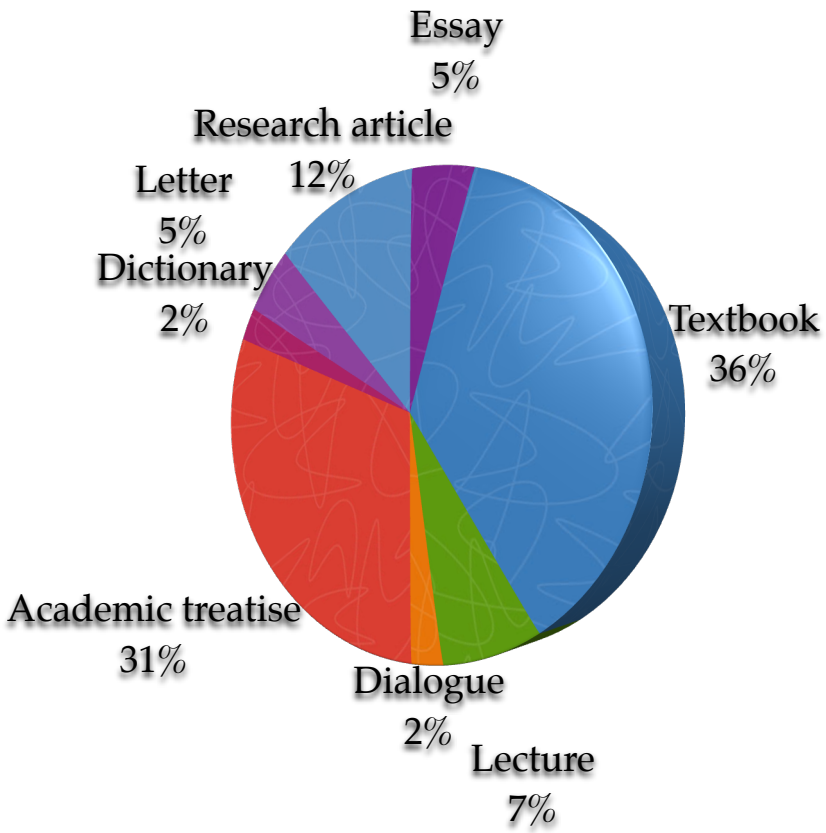


Figure 11: *Genre/text-type of the sample texts in CETA.*

The distribution between more formal texts –treatises, articles and essays– and texts aimed at a non-specialized audience is rather balanced –45% versus 55%–. This balance facilitates the inferring of information using this text-type variable. With fifteen sample texts, textbooks are the most common category, followed by academic treatises, which make up to twelve texts. Research articles are slightly less frequent and they are mostly found in the second half of the nineteenth century. The didactic character of the three lectures in *CETA* is already visible in their titles[67]. Essays and letters are present in the corpus with two instances each and there is also a sample from Hill's (1754) comprehensive dictionary on astronomical terms. Finally, Harris' *Astronomical dialogues between a gentleman and a lady*, published in 1719, is the only example of dialogue.

Moskowich (2012, p. 41) remarked that "these categories were all in use at the time when *CC* authors were producing their texts and many of them have been recorded as early as the end of the fourteenth century." It is true as well that although these text-types existed, there may be some minor discrepancies in their conception by both authors and audience today and two centuries ago. However, all in all, these slight differences do not minimize the importance of these texts as long-established reference models for science writing. Moskowich (2012, p. 41) highlighted that the plurality of text-types in *CETA* is a consequence of the compilers' emphasis on reflecting the social reality of astronomy and the writing of science in the period.

Regarding sex of authors, there are only two women in the corpus, which represents only 5% of the total. This lack of balance, however, is not due to a male-

[67] Whiston's (1715) *Astronomical lectures*, read in the publick schools at Cambridge, Ewing´s (1809) *A plain elementary and practical system of natural experimental philosophy: including Astronomy and chronology*, and Phillips' (1817) *Eight familiar lectures on Astronomy: intended as an introduction to the science: for the use of young persons and others not conversant with the mathematics.*

oriented text selection but rather to historical accuracy. The number of women doing science, writing and publishing under their true name texts about astronomy in the eighteenth and nineteenth centuries was indeed very small, as discussed in chapter one.

The first female author in the corpus is Margaret Bryan (1760-1815). There is evidence that she married a schoolmaster and ran schools at various times in London and Kent. She taught astronomy and natural philosophy to girls at a time when education for ladies was largely non-institutionalized. She also wrote elementary, practical books about general astronomy and physics. Her first work, published in 1797, was the *Compendious System of Astronomy*, a collection of notes and lectures in book form. Its success encouraged her to publish two more books: *Lectures on Natural Philosophy* in 1806, and in 1815, *Astronomical and geographical class book for schools.* The anonymous *Conversations on Chemistry* (1806) is often also attributed to her. All her books were conceived of as instructional texts for children. *CETA* contains a sample from Lectures IV, V and VI from the *Compendious system of Astronomy*. Apart from the 1797 edition, other two editions were published, which shows that Bryan's work was well received by the public.

The second female writer, Agnes Mary Clerke (1842-1907), was born at Skibbereen (Cork, Ireland). Clerke's active years spanned a period of enormous development in astronomy. She lived through the last few years of pre-Einstein science, became the leading commentator on astronomy and astrophysics in the English-speaking world, and was said to be the most influential writer on astronomy in English language at the end of the Victorian era. As mentioned in chapter one, Clerke received a home education from her father and her brother, both scholars at Trinity College. As a child she became interested in astronomy due to her father's influence. She began to

write about the history of astronomy before she was fifteen. In 1861 the family moved to Dublin, and in 1867 she and her sister travelled to Italy, where they stayed for a decade. She studied in Florentine libraries and wrote her first article on astronomy, "Copernicus in Italy", published in the *Edinburgh Review* in July 1877. A *Popular History of Astronomy in the Nineteenth Century* (1885) was her first book. Five years later she published *The System of the Stars*. In 1892 the Royal Institution awarded her with the Actonian Prize. She was also a member of the Astronomical Association, and was elected honorary member of the Royal Astronomical Society, together with Lady Huggins[68]. Clerke's third book, *Problems in Astrophysics*, was published in 1903. Her work gathered both a literary and a scientific reputation. As a mature writer, she also produced several short volumes and she contributed articles on 150 astronomers to the *Dictionary of national biography*, and other encyclopedias, like the *Encyclopaedia Britannica*, and the *Catholic Encyclopaedia*. She also contributed widely to periodicals such as *Knowledge*, *The Observatory*, and *the London Tablet,* among others.

These two women included in *CETA* are, in some way, not the most repeated pattern for women scientists in the eighteenth and nineteenth centuries. As explained in chapter one, there were two main profiles for women concerned with science in these centuries: the noblewoman attending salons and informal scientific discussions and the invisible assistant, that is a woman helping her male relative in his scientific activities. Both Bryan and Clerke surpassed the expectations of their time, did science by themselves and publish with their own name.

[68] This was indeed a great recognition, as only three other women, Caroline Herschel (sister of the well-known John Herschel), Ann Sheepshanks and Mary Somerville shared this honor.

Concerning occupation of authors, figure 12 shows the distribution of professions of the authors in *CETA*. As could be expected, the highest percentage corresponds to professors and other academic staff working at universities.

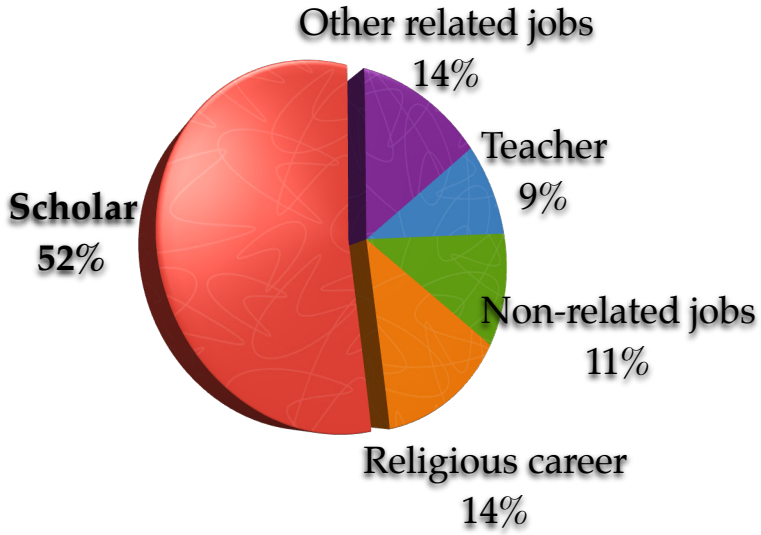


Figure 12: *Occupation of authors in CETA.*

In *CETA* there are 14 writers that were appointed professors at leading universities: William Whiston (1667-1752), Roger Long (1680-1770), Matthew Stewart (1717-1785), John Ewing (1732-1802), William Herschel (1792-1871), Landon Garland (1810-1895), Denison Olmsted (1791-1859), William Bartlett (1804-1893), George Darwin (1845-1912), Charles Young (1834-1908), Ormsby Mitchel (1809-1862), William Whewell (1794-1866), Elias Loomis (1811-1889), William Chauvenet (1820-1870), and Percival Lowell (1855-1916). Within the group of professors and

scholars, those who were not appointed professors had a wide range of occupations: John Harris (1866-1719) and James Hodgson (1672-1755) were Fellows of the Royal Society, James Ferguson (1710-1776) and John Bonnycastle (1750-1821) were lecturers and the rest (George Costard (1710-1782), Samuel Vince (1749-1821), and David Brewster (1781-1868) held other university positions. The number of astronomers working in education at non-university levels (four) is lower but still remarkable. In most cases, these teachers ran schools in more rural settings, like John Gummere (1784-1845).

The influence of religion on science was still great at this period, as can be seen in the high number of writers (six) that developed their career within the Church. The pattern of authors combining scientific and religious careers is repetitive. George Costard (1710-1782) combined his scientific interests working as a proctor at Oxford University in his youth, and then he became Vicar of Twickenham. Similarly, Samuel Vince (1749-1821) worked as an astronomer at Cambridge University and was also a clergyman whereas Robert Small (1732-1808) was ordered minister of the Church of Scotland and appointed Fellow of the Royal Society of Edinburgh. The only American writer serving the Church was John Ewing (1732-1802), who was the Pastor of the First Presbyterian Church of Philadelphia from 1759, one year after he accepted a position as Professor of Ethics at the College of Philadelphia.

The category of "other related jobs" includes those authors working on areas closely related to astronomy (six) that were not concerned with education and research at universities: Robert Morden (1668-1703) and George Adams (1709-1772) were instrument makers and William Nicholson (1753-1815) was a shipman at the service of the East India Company. Concerned with the world of editing and publishing was

William Philips (1775-1828), who entered his father's business as a printer and Agnes Clerke (1842-1907), a reputed writer on science.

Finally, five other *CETA* authors developed their careers in fields not traditionally associated with astronomy. This is the case of Henry Curson[69], who worked as a lawyer in London, Jasper Charlton, who was an officer of the customs at Gainsborough, and John Hill (1716-1775) who owned a pharmacy in St Martin's Lane, Westminster. Clearly enough, the multifarious nature of the careers pursued by authors in *CETA* proves that three centuries ago, the path for making science was more open than today.

The geographical variable reflects not only the actual place of birth of authors but also the place where they spent their childhood and acquired their linguistic habits. In theory the corpus could contain any text written in any of the (former) colonies of the British Empire. However, in practice, only texts from English, Scottish, Irish and American authors have been included in *CETA*[70]. The compilers of *CC* decided to create four main categories for the parameter of place of education, namely England, Scotland, Ireland and the U.S. These are historical nations rather than political nations. At the time the texts in the *CC* were published, the only official nation existing was the United Kingdom. Only Young's *Recent progress in solar astronomy*, published in 1880 and Lowell's work on Mars from 1895 were published after the *American Declaration of Independence*. In the case of Irish writers, all the works written by Irish authors were published when Ireland was still a colony of the U.K. Figure 13 reflects the percentages of the places of birth of the authors in the corpus:

[69] Curson's and Charlton's birth and death dates are unknown.

[70] *CHET*, the history subcorpus, also contains some Canadian authors.

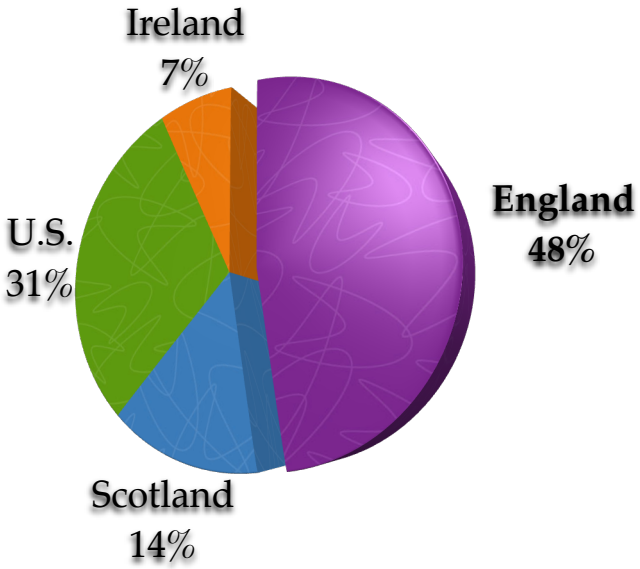


Figure 13: *Place of education of authors in CETA.*

The apparent unbalanced nature of all the categories in this figure reflects indeed the historical development and situation of science in English-speaking countries in the eighteenth and nineteenth centuries. The fact that 50% of the samples, that is, twenty-one texts, are signed by English authors is a reflection of history: during the eighteenth century science was almost exclusively an English enterprise. As explained in the first chapter, during the eighteenth century the American colonies were still dependent of England, and that included science as well. The independentist ideas that came with the nineteenth century fostered the development of American science as well. Indeed, the first text contained in this corpus and written by an American author is Ewing's 1809 *Lecture on natural experimental philosophy* and the total of texts written by Americans amounts twelve. On the other hand, the six texts authored by Scottish scientists (Brewster, 1811; Croll, 1889; Ferguson, 1756; Small, 1804; Stewart, 1761; Wilson, 1774) can be said to be a consequence of the Scottish Enlightenment that took place in

the eighteenth century and that multiplied the number and quality of artistic and scientific activities in Scotland, partly as a nationalistic reaction to the *Act of Union* (1707) that joined England and Scotland together[71].

Apart from the nation in which the authors of the corpus were educated, the metadata files in the *CC* also offer information about the specific regions and cities of education. Unfortunately this information is not available for all of them and, whereas in some cases compilers have been able to track down the author's education and include the places of primary, secondary and college education, in other cases all kind of personal information about the author is missing. Figure 14 shows the distribution of the place of education for European authors:

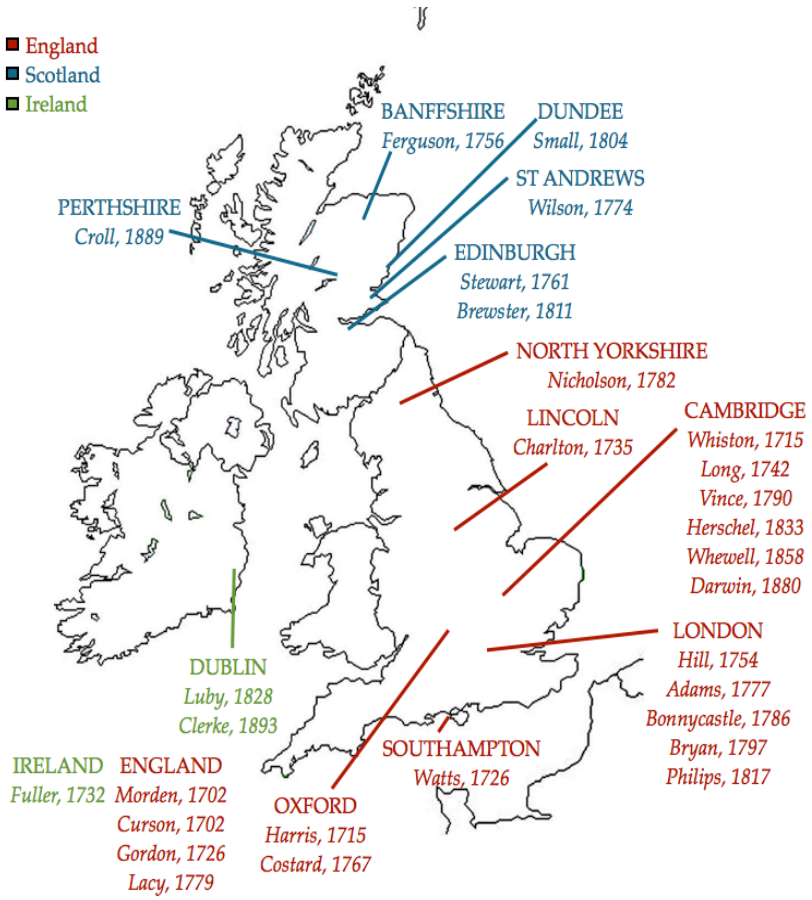


Figure 14: *Distribution of European writers in CETA.*

[71] See section 1.2.

This figure contains twenty-eight of the thirty European writers contained in *CETA*. According to the metadata in *CETA*, George Gordon was European but no further specification about his place of education is given and that is why he has been left out of this map. Similarly, this map does not respect the categories presented in tables 4 and 5. As stated before, the idea was that of erasing the category "unknown" in the geographical variable because, even if it is more accurate, it would eliminate texts from analysis. As a result, for data analysis, I will consider author´s nationalities as stated in figure 14, that is, Curson (1702), Charlton (1735), Hodgson (1749) and Lacy (1779) will be considered English and Fuller (1732), Irish. George Gordon (1726) will be considered European and taken into account only in the "Europe/America" analysis.

As can be seen in the map, the great majority of European writers were English. Within this country, 65% of writers were educated in the region of the invisible college, where the leading universities and scientists were placed in England. As discussed in chapter one, the allegiances of the scholars from the universities of Cambridge and Oxford and Gresham college during the seventeenth century resulted in the foundation of the Royal Society and motivated the consideration of this area as the center of scientific and scholarly activities in England, a consideration that is still running nowadays. As a result of this process, the apparent unbalanced geographical distribution of the texts in *CETA* reveals itself extremely accurate. Concerning non-English authors, Alexander Wilson (1714-1786) received his Master´s degree from St Andrews university in 1733 and Sir David Brewster (1781-1868) was sent to the University of Edinburgh when he was twelve was awarded an Honorary Master of Arts degree when he was nineteen.

The foundation of the Royal Society was an ongoing process at the time texts of *CETA* were published and, as a logical consequence, some of these authors were related in some way or another to the academicist movement and its leading figures. John Hill (1716-1775), for instance, unfortunately tried to become a Fellow of the Royal Society but his friendship with Martin Folkes, Sir Hans Sloane, Henry Baker, William Watson, and James Parsons, members of the Society, is credited. Others were more successful and entered the Society. John Herschel (1792-1871), son of the famous astronomer William Herschel was educated at Cambridge's St. John College and after graduation, he was elected Fellow at this college and at the Royal Society, where he started a long career of advancements in astronomy and chemistry[72]. Outside England, Robert Small (1732-1808) was elected as a member of the Royal Society of Edinburgh on November 11, 1783.

Some of these authors were not only students at these leading universities. William Whiston (1667-1752), for example studied at Clare College at Cambridge and worked for great part of his career as Senior Fellow there with Isaac Newton. Similarly, Roger Long (1680-1770) was also a teaching Fellow at Cambridge's Pembroke College as well as a Fellow of the Royal Society, Reverend John Harris (1666-1719) was a scholar for a short period at Trinity College Oxford. In Scotland, Matthew Stewart (1717-1785) was elected Professor of Mathematics at Edinburgh University in 1747. Other professors of mathematics and astronomy at Cambridge were Samuel Vince (1749-1821) and George Darwin (1845-1912), son of the famous biologist Charles Darwin.

[72] He is believed to have set the scientific basis of photography.

Regarding American authors, the second more populous group in *CETA*, in the nineteenth century only the most advanced cities on the East coast were in the situation of producing good scientific work. As explained, in chapter one, American colonists were largely dependent from London and their interests were not always directed at the advancement of astronomy. Figure 15 shows the distribution of American authors in *CETA*. As in the case of European writers, this will be the distribution used in data analysis. Writers have been listed in four groups (New England, Midwest, Mid-Atlantic and South Atlantic) according to their place of education and the judicial system of the country. *CETA* compilers did not make geographical differentiations within American authors, as they did with Europeans and that is why this study will take a slightly different approach in the geographical variable. Hence, from the four groups suggested by *CETA* compilers (England, Ireland, Scotland, and the U.S.), in this work data analysis will be made up of seven groups. First, analysis according to geographical distribution will focus on continental differences, that is a contrastive "Europe versus America" analysis. Both continents will be dealt with separately. In the case of European texts the three groups used will be the historical nations of England, Scotland and Ireland[73]. Concerning American texts, they will be classified in the four groups showed in figure (15). These groups coincide with present-day U.S. judicial system and in popular culture could be said to be an equivalent of European *historical nations*[74].

[73]As mentioned before, it could be argued that Ireland is more than a historical nation. Leaving aside all political discussion, the creation of the Republic of Ireland took place after the time-span covered by *CETA*.

[74] New England or Midwest are commonly used as pseudo-nationalistic denominations.

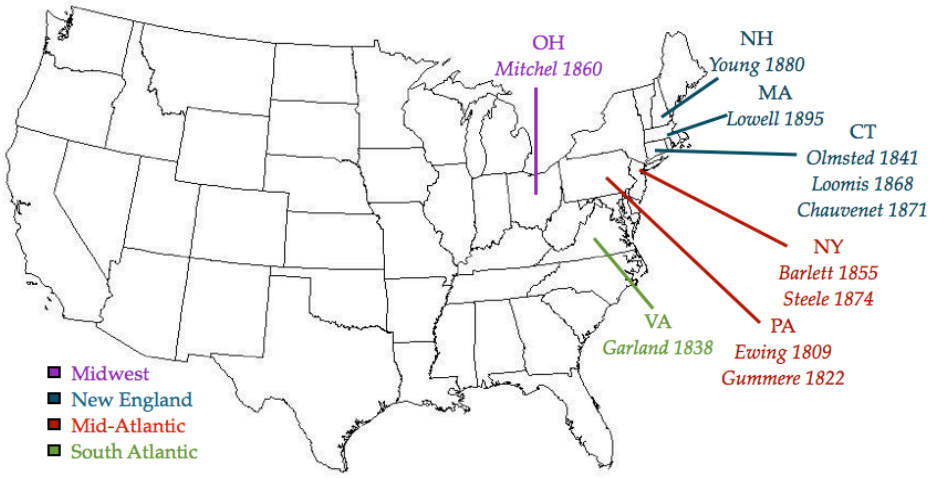


Figure 15: *Distribution of American writers in CETA.*

Historically, the region of New England (marked with blue) had worked as the motor of the country. For this reason, it is not surprising that most writers were educated in the states of Massachusetts, New Hampshire or Connecticut. New England has the oldest universities and has always been the academic center of the country. Half of the New Englanders included in *CETA* received college education at one of the Ivy League universities: Charles Young (1834-1908) attended and was professor at Dartmouth College, Percival Lowell (1855-1916) studied at Harvard and Elias Loomis (1811-1889) graduated from Yale University.

Even if New England was politically leading the country, Philadelphia (Mid-Atlantic region, marked in red) was known in the nineteenth century as the scientific capital. The American Philosophical Society was founded there. As a consequence, the

Mid-Atlantic States of New York and Philadelphia witnessed a rapid growth of scientific facilities which made it easier for scientists to carry out their work. *CETA* writers from this area also attended prestigious institutions and, for instance, William Barlett (1804-1893) graduated from West Point Academy. The case of John Ewing (1732-1802) is indeed quite representative of the situation of scientists in America. He graduated from the College of New Jersey in 1754 and then, after several years of traveling, received his doctoral degree from the University of Edinburgh. As Greene (1984) remarked, it was rather normal that American scholars would complete their education in Europe, as institutions there enjoyed great prestige. Finally, *CETA* also has one author representing the states of Virginia (Garland, 1838) and Ohio (Mitchel, 1860).

According to the kind of education authors received, it is clear, as figure 16 shows, that the great majority received a university education. Indeed, most of them were professors at leading universities. However, university was not the only way of building a scientific career in the eighteenth and nineteenth centuries. Unlike today, when university is the only accepted way to enter the scientific career, in those centuries attitudes to learning were more flexible. Thus, we find examples like that of Isaac Watts (1674-1748), who received his education at John Pinhorne's grammar school and later acquired his scientific knowledge while working as a tutor for the family of Sir John Hartopp at Stoke Newington. Similarly, John Gummere (1784-1845) received his formal education in country schools and later taught himself about algebra, astronomy, geometry and trigonometry as preparation for his classes in several boarding schools.

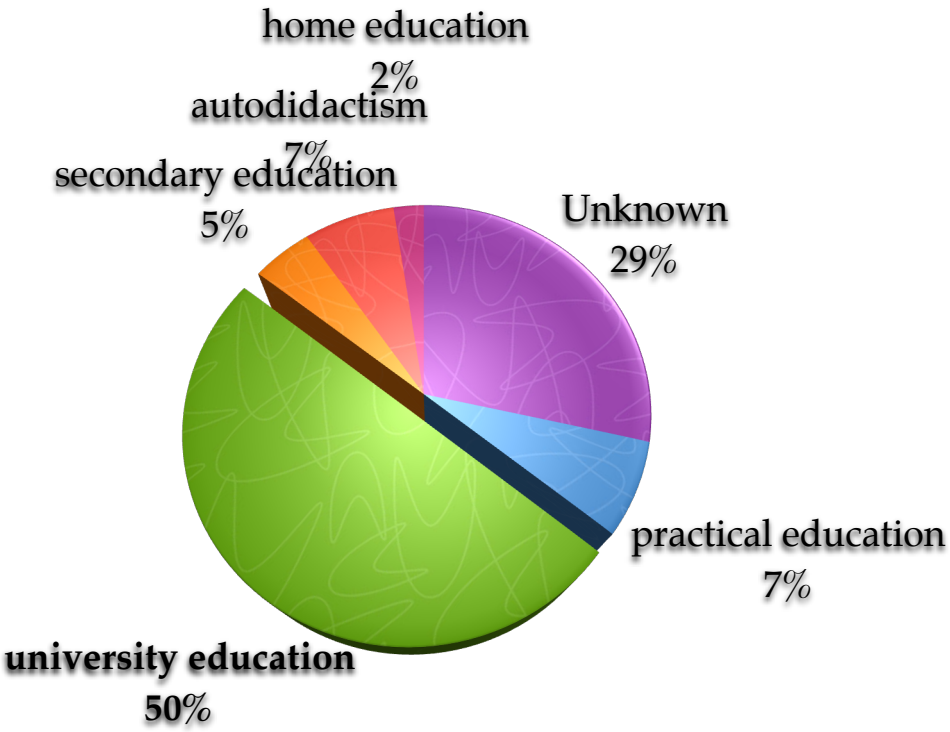


Figure 16: *Education of authors in CETA.*

Another option was to learn science through the execution of a trade (practical education, 7%), a precedent to today's applied branches of sciences. Robert Morden (1668-1703) was a reputed map seller, cartographer, globe and instrument maker, publisher and bookseller. His *Introduction to Astronomy, Geography, Navigation, and other Mathematical Sciences Made Easie by the Description and Uses of the Celestial and Terrestrial Globes* (1702) was consequence of his skill as a cartographer but science was not his only concern. In 1672 he realized that the number of counties in England and Wales (52) coincided with the number of cards in a pack. He thus had the idea of publishing playing cards which illustrated each county. The cards sold very well and

other publishers soon realized that the middle classes, while wary of frivolous card playing, would buy such educational packs. John Hill (1716-1775) worked as an apprenticeship with an apothecary and it was his career and his friendship with some Fellows of the Royal Society that triggered his interest in natural science. Another common profile of the time was that of William Nicholson (1753-1815) a member of the East India Company, who acquired his knowledge about navigation and practical astronomy sailing to the East Indies.

The group of the self-educated scientists (7%) is also present in *CETA*. James Ferguson (1710-1776) learned to read by listening to his father teach his elder brother the catechism. His interest in astronomy led him to invent a cardboard instrument for showing the motions and positions of the sun and moon. Ferguson later was elected a Fellow of the Royal Society. A similar situation was that of John Bonnycastle (1750-1821), who, coming from a modest family and lacking the advantage of a classical education, was appointed Mathematical Master and Lecturer in geography and philosophy at the Royal Military Academy in Woolwich. Worthy of mention, the case of Agnes Mary Clerke (1842-1907) can be considered representative of the great majority of women interested in science in this period. Clerke´s father was a classical scholar and graduate of Trinity College, Dublin, and an amateur astronomer. Her only brother, Aubrey St John Clerke (1843-1923), was a brilliant scholar and medallist in mathematics and science at Trinity College, Dublin and later a Chancery barrister in London. As many women, both Agnes and her sister Ellen had the chance of accessing science thanks to a learned open-minded family and cultivated their interest at home. Clerke's case was quite unusual in the sense that she not only was allowed to pursue her

scientific interests but also she had the chance of publishing more than fifty articles, mainly on scientific themes, in the *Edinburgh Review*.

This sociological review of *CETA* authors has served again to contextualize study texts. After considering issues like sex of authors, occupation, place and degree of education it becomes clear that authors respond, in one way or another, to the historic situation portrayed in chapter one. This is the best argument to prove that *CETA* is indeed a representative corpus because it represents the variability of the population of science writers in English-speaking countries in the eighteenth and nineteenth centuries to its whole extent.

3.1.2.2. Parts of *CETA*

Together with the electronic version of the samples of the scientific texts compiled for the CC, there are a series of files containing sociolinguistic information about each of the texts and authors in the corpus. The use of metadata files is especially helpful for the extraction of sociolinguistic information for interpretation together with purely linguistic features. These series of metadata files can be consulted through the *CCT* to widen the information about texts and author and the information contained in them can also be used as a parameter for advanced searches. Thus, for instance, the *CCT* allows the possibility of restraining our query only to texts written by American authors or academic treatises.

In figure 17 the two main categories in which metadata files are divided can be clearly seen. The first part of the file (marked with arrow 1) is concerned with information about the author. Specific data like sex, birth and death years, occupation and place of education are included, together with a brief biography in the form of narrative. The second part of the file (arrow 2) provides information about text: text-type, year and place of edition, complete title and some brief information about the texts, which may include the number of subsequent editions that were produced, the belonging of the book to a series of books, its reception by the audience and other considerations of the kind.

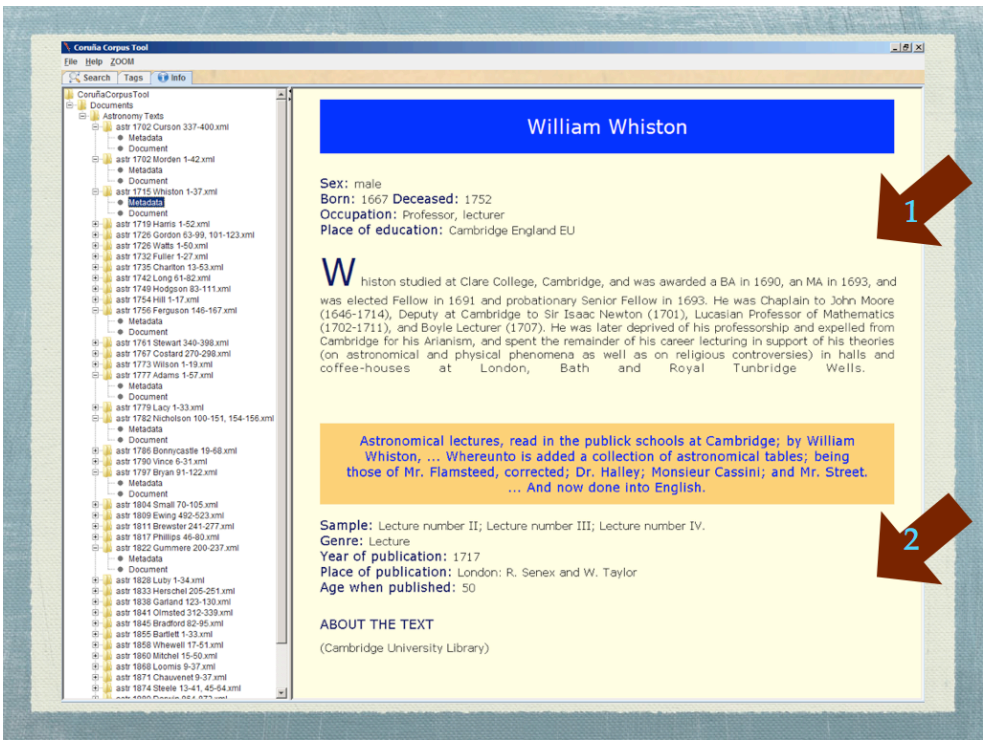
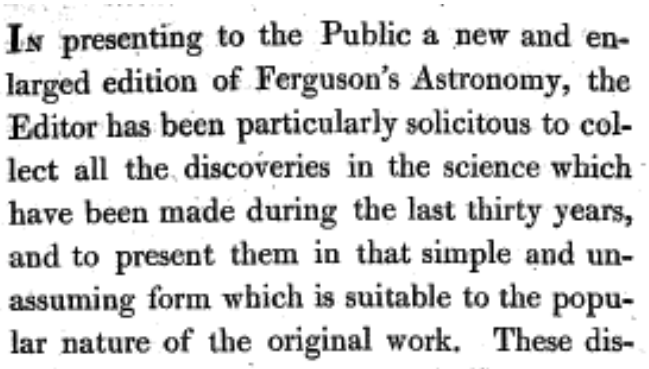


Figure 17: *View of metadata on the CCT.*

Prefaces are not contained in the corpus and cannot be accessed by the general public but, they are a great source of information for understanding the motivations, desires and mental configurations of writers in relation to their pieces of work. This was the decisive reason for the inclusion of a short study of them in this piece of work. Prefaces are historical documents, whose reading can provide first-hand insight on how authors understand their writing process. This information is precious, especially in historical linguistics, where access to information about texts is very scarce. Obviously, not all the texts in *CETA* contain a preface. Among those texts with a preface, several concerns are repeatedly shared by writers. These concerns range from the explanation on choices about the contents of the work, normally reasoned on text-type constraints, to motivations for writing. The most repeated feature in prefaces is the insistence on the acquisition of a plain style, which can be read as an attempt to follow Bacon's indications on how to write experimental essays[75]. Indeed, in many cases, these allusions to simplicity are highly ambiguous and they are rather rhetorical than real.

Some curiosities can also be pointed out, such as, the existence of network connections among writers. In the following quotation, Brewster (1811) made explicit reference to what is already obvious from the title –Ferguson's *Astronomy explained upon Sir Isaac Newton's Principle*– Brewster's work is based upon Ferguson's *Astronomy Explained upon Sir Isaac Newton's Principles and Made Easy to Those who Have not Studied Mathematics*, another academic treatise published in 1756 and also present in *CETA*:

[75] See section 1.3.1.



In presenting to the Public a new and enlarged edition of Ferguson's Astronomy, the Editor has been particularly solicitous to collect all the discoveries in the science which have been made during the last thirty years, and to present them in that simple and unassuming form which is suitable to the popular nature of the original work. These dis-

This quotation may serve as a reminder that texts and authors are not individual entities; they belong to groups and therefore a certain degree of group codification must be expected.

Texts constitute a special way of communication, and prefaces play an important role in this communication process. Following Jakobson's (1981) teachings about the functions of language and the process of communication, it is possible to include prefaces as a special type of contact between the sender and the receiver of the message. Figure 18 illustrates how Jakobson's schema for the process of communication could be adapted to *CETA* texts. For this particular figure, the particular case of John Bonnycastle's *An Introduction to Astronomy, in a Series of Letters from a Preceptor to his Pupil* from 1786 has been chosen.

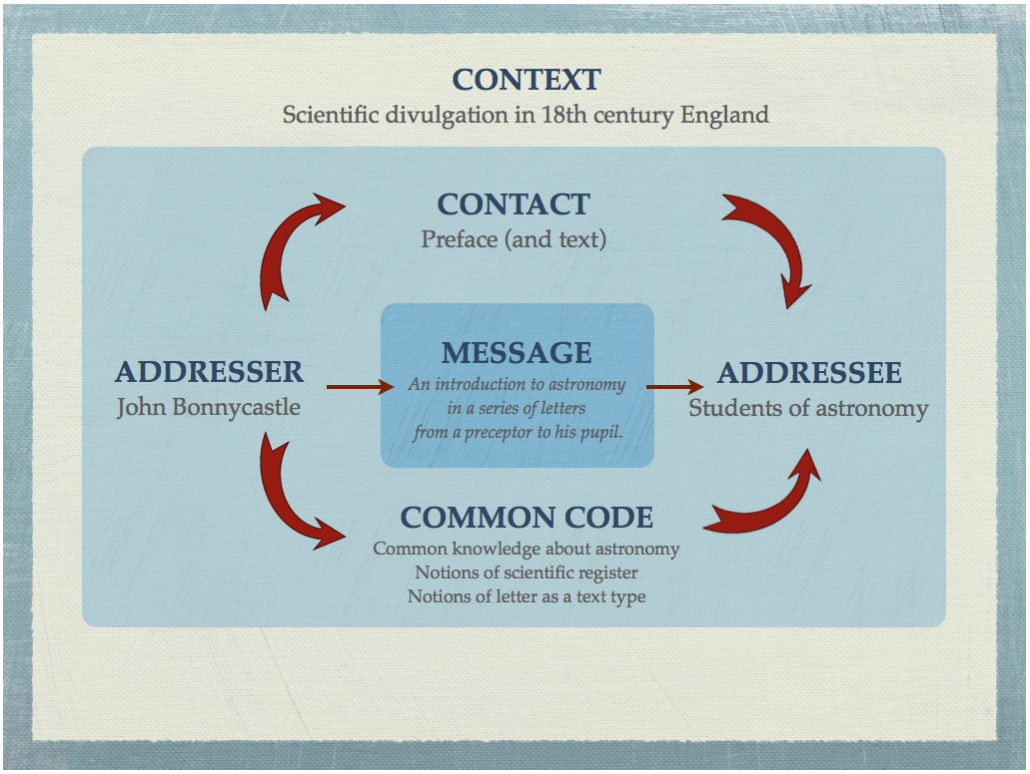


Figure 18: *The process of communication in CETA.*

Prefaces can be listed as a way of contact. For texts without a preface, the category "contact" could also be filled by the text itself but prefaces constitute the best opportunity for authors to address directly to readers, as they are normally written in first or third person singular (*I* or *the author*) and addressed directly to the reader. Additionally, prefaces usually contain specifications about the intended addressees, as well as the aims and motivations of the work. The role of prefaces, however, goes far beyond appearances. Without a preface, only the categories of "author", "message" and "context" could be inequivously described. Prefaces are especially useful in describing the intended addressees of the message, as can be seen in the following quotation, extracted from Bonnycastle's preface (1786, p. iii):

Many, who are not ſufficiently acquainted with the Mathematics, to read, with ſatisfaction, the works of Newton, and other eminent writers upon this ſubject, are yet very deſirous of obtaining ſuch an idea of it, as will enable them to comprehend the leading principles upon which it is founded, and to partake of thoſe pleaſures, which enquiries into Nature, and the inveſtigation of ſome of her grandeſt operations, muſt neceſſarily afford to every ingenuous and inquiſitive mind.

To this claſs of readers, the following Letters are particularly addreſſed.

In this quotation, Bonnycastle made it clear that his work is aimed at a non-specialized audience. This piece of information had already been stated in the title (*An introduction to Astronomy, in a series of letters from a preceptor to his pupil*) but the preface extends the information and specifies how it should be understood. Readings of the text by other people outside the group specified by the author –other members of the scientific community or the general public– may, of course, be possible. In those cases, addressee specifications could be taken as reliable parameters for the creation of a realistic expectation horizon in readers. Concerning addressees, there are multiple allusions to the suitability of texts to students, as can be seen in the following quotation (Costard, 1767, p. viii):

The Author preſumes his Work to be the firſt of the kind, and hopes it will prove beneficial to Students, for whom it was chiefly intended.

In this case, Costard explained that this work is a textbook written for students. A similar message can be found in the comment made by J.T. Desaguliers, the editor of Gordon's *Introduction to Geography, Astronomy, and Dialling* (1726):

Jan. 11. 1725-26.

I *Have read this Book, and think it will be very uſeful to Beginners.*

J. T. Deſaguliers,
L. L. D. F. R. S.

The comment made by the editor slightly contradicts text-type classification. Academic treatises are supposed to be advanced texts for learned readers. However, the editor specifies that this text can be useful for learners. This type of incongruence is consistently repeated throughout all the prefaces studied, which indicates that plainness in style was indeed a desired feature for scientific writings.

On the other side, prefaces may also reveal much information about the common code that addressers and addressees should share for an effective communication. In prefaces, writers delimit the topic of the book, set minimum requirements for readers and provide information about their understanding of the topic, so that readers can understand and share the code authors have used to produce their text. An example of this can be found in the following quotation taken from Gummere's *Elementary Treatise in Astronomy* (1822: iii)

In the demonstrations, the student is supposed to be acquainted with Algebra, Geometry, Plane and Spherical Trigonometry, and Conic Sections, or at least the properties of the Ellipse.

Gummere addressed his work to students but he set a minimum of content requirements. These specifications are, however, not related to the style or the language

used, but to the contents included in the text. Following the didactic purpose of textbooks like this, these specifications may be understood as fulfilling a metalingual function[76]; the author (or teacher) is trying to verify that the receiver (the student) *speaks* the same language he does. The student's inability to comply with the content requirements stated here would result in a failure in communication. The author's ultimate goal is therefore explaining his code with the expectation that the reader will share it. A similar quotation can be found in Stewart's preface to his *Tracts* (1761, p. vii)

> *As the reader is ſuppoſed to be well acquainted with Euclid's Elements, there are no references made to them*

This quotation by Stewart is not directly about language but it can be deduced that knowledge requirements in a text may result in language requirements as well. Unlike Gummere's elementary treatise, Stewart's *Tracts* constitute a learned text aimed at specialized readers. Indeed, this is one of the turning points of text-typing and scientific register and one of the hypothesis of this study is to prove that text complexity –in this case, the use of nominalizations– has been favored in English scientific register due to guild codification and the development of a specific register used only for and by specialists that remains alien to a non-specialized audience.

Another important conclusion that can be easily inferred after reading all the prefaces is that text-types were indeed known to writers at this period. It may be argued that the boundaries between lectures, textbooks, treatises and scientific articles may have shifted slightly in the last 200 years but there is evidence proving that scientists in

[76] Jakobson (1981) defined metalingual function as those elements that make sure addresser and addressee understand each other.

the eighteenth century understood and used text-types in a similar way than we do today. The following quotation (Hill, 1754, p. 8) may serve as an illustration of this assertion.

all that is inſtructive, and all that can be entertaining in it, is laid down in the fulleſt manner; and as the book is intended for the uſe of thoſe who are not accuſtomed to calculations, the eye is not offended with the figures.

Had this work been written for aſtronomers, a very different method had been purſued;

In this quotation, Hill stated clearly that his way of writing his dictionary would have been different if the work would not be aimed at beginners. Although he only mentioned explicitly the use of figures, it could be assumed that other features, like the language used, were taken into account to simplify the reading. The specification of the contents and the linguistic choices made is indeed one of the reasons why prefaces are so important for text-type assignation. Other than in prefaces, information about text-types can be found in the title and in actual texts. This last option is highly problematic because, dealing with historical texts, text-type considerations are sensible to have suffered slight modifications. For this reason, comments made about the authors of the texts in prefaces are especially valuable. In some cases, text-type is delimited and explained by the author, as in the case of Phillips' preface (1817, p. ii)[77]:

[77] It may be noted here that the title of Phillips' work already contains the word *lecture*: *Eight familiar lectures on astronomy: intended as an introduction to the science: for the use of young persons and others not conversant with the mathematics*. This clearly facilitates text typing although the detailed explanation found in the preface can definitely work as a confirmation.

I repeat that, in undertaking these Lectures, the hope of being useful was my only object. And, being aware from past experience,* that a large proportion of those who might be expected to attend them, would consist of young persons, many of whom were altogether unacquainted with the science; it became my anxious determination to render them as simple as possible, as the only means of their being advantageous to that class of my hearers.

When referring to the style used in his book, Phillips (1817, p. iii) explained that he adopted a simple style that is more adequate for students, as he concluded from his own experience:

a regard was had to plainness of language, which would necessarily have been tedious to the scientific; but which is, as I conceive, highly advantageous to the learner: who also will more readily excuse the occasional or even frequent repetition that will be found in the following pages.

In this quotation, Phillip addressed directly the issue of language in scientific register. His work is mainly intended for students and as a result he adopted a plain style full of repetitions, more appropriate for a pedagogical text. His attitude in excusing his style for being tedious for advanced readers proves the claim that authors in this period were indeed aware of text-type distinctions and had them in mind when writing their works. Concerning linguistic choices made explicit in prefaces, writers seem to be fixed with the idea that a good scientific piece of work must be plain both in contents and style. This is a direct consequence of Francis Bacon's influence. In his *Essayes*, originally published in 1696, Bacon (1985, p. 105) praised simplicity in discourse by saying that "[s]peach of a Mans Selfe ought to be seldome, and well chosen." Instead of presenting their studies in a lengthy, comprehensive way, authors ascribed consciously

to Bacon's standard as can be seen in the following quotation, from Bonnycastle's letters (1786, p. iv):

> The principal object in view, throughout the whole performance, has been to avoid, as much as poſſible, all complicated mathematical principles and calculations, and to elucidate the moſt ſtriking particulars, in as popular and eaſy a manner as the nature of the ſubject would admit.

This quest for simplicity is indeed a common denominator in all text-types. In some cases, the specifications found in prefaces are to be understood with caution, as the boundaries between real description of texts and rhetorics are far from clear. Such is the case with Luby´s treatise from 1828. The title of the work, *Introductory treatise on Physical Astronomy*, sets our horizon of expectation in the field of formal texts. The reading of the preface, however, renders this classification dubious, as it is claimed that the main addressees of this treatise are initial students of physical astronomy. Luby was a lecturer at Trinity College and therefore acquainted with the use of formal texts. His consideration of a learner's text and his willingness to comply with Bacon's standards of simplicity in texts are issues beyond the reach of this work. However, from this example and from the study of all the prefaces of the texts in *CETA*, it can be concluded that even if text-type considerations were known to writers in this period, the omnipresent desire for plainness in style is sometimes a rhetorical strategy writers used to comply with the standards for scientific method and style established by Bacon.

Format is one of the main features of modern corpora, as all texts are electronically saved and transmitted. Indeed, the rapid advancement of computer science has often been pointed out as one of the main reasons for the establishment of

the discipline of corpus linguistics in the last decades of the twentieth century (Hickey, 2003). In the case of historical texts, the treatment, edition and creation of electronic corpora constitute an additional challenge. After text selection, sample texts in *CETA* followed a conversion process from original paper publications to electronic XML files, as exemplified in figure 19.

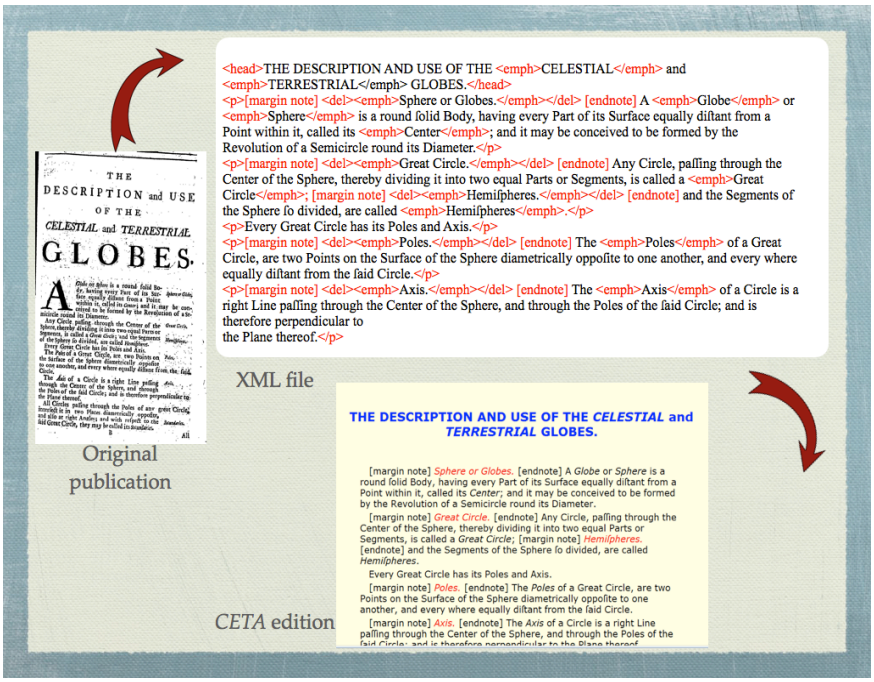


Figure 19: *Conversion process for CETA sample texts*.

At the earliest stage, the selected original publications were scanned and saved in a PDF file. Then, those files were converted into text files using optical character recognition (OCR) software whenever possible. In the case of early texts this stage was rather problematic due to the existence of old-fashioned characters such as the ligature <ct> and long <s>, represented as <ſ > and its italicized form <ſ>. In cases with heavy profusion of these characters, which were not recognized by OCR[78] software, manual typing and extensive revision were needed. The same problem was found in the case of

[78] Optical Character Recognition.

special astronomical symbols of formulae, special features of astronomy texts.Once texts were converted into a text file and several revisions were made to ensure good representation standards, *CETA* sample texts were encoded following the Text Encoding Initiative (TEI), an international project whose main aim is to provide standardized implementations for the interchange of machine-readable texts (McEnery & Wilson 1996, p. 23; Crespo, 2011)[79]. Texts compliant with TEI are made up of two parts, namely the "HEADER" and the "TEXT". The header provides information about the author, the title and date of edition of the machine-readable text, as well as other information about encoding decisions adopted. TEI-compliant texts use Standard Generalized Markup Language (SGML) for producing human and machine-readable texts. In the case on *CETA*, the markup language chosen was the Extensive Markup Language (XML), which is one type of SGML document markup worldwide recognized for its clarity, simplicity and rigurosity.

CC is an annotated corpus and, accordingly, in it a set of tags is used to mark specific elements in the text. These tags are not visible in the final stage of the text but are present in the XML file. Tags are easily recognized because they are contained between two pairs of angled brackets. The start tag appears at the beginning of each element and contains an annotation string (for instance <emph> the start tag marks a string of words with a special emphasis, either italization or boldness). Then, the end tag repeats the same structure of the start tag but with the addition of a slash before the annotation string (</emph>).

[79] Further information can be found on the TEI's webpage http://www.tei-c.org/index.xml (Retrieved October 15, 2012).

Figure 20 shows the final view of an encoded *CETA* sample text on the *CCT*. Arrows two to four mark specific points within the "HEADER", and arrows five and six signal the "TEXT" itself. The main advantages of an annotated corpus are obvious at first sight because, as McEnery & Wilson (1996: 24) pointed out, the information that was "implicit in the plain text has been made explicit through concrete annotation."

Arrow one signals the tab within the *CCT* search engine where one can access the view of texts and metadata files. Arrow two marks the location of the title of the file (astr1782Nicholson.xml, in this case), information about the editor of the machine-readable text (Isabel Moskowich-Spiegel), as well as the research group (MuStE) and the entities providing funding for the project (Xunta de Galicia, Universidade da Coruña and Deputación da Coruña). This information is present in all *CC* files.

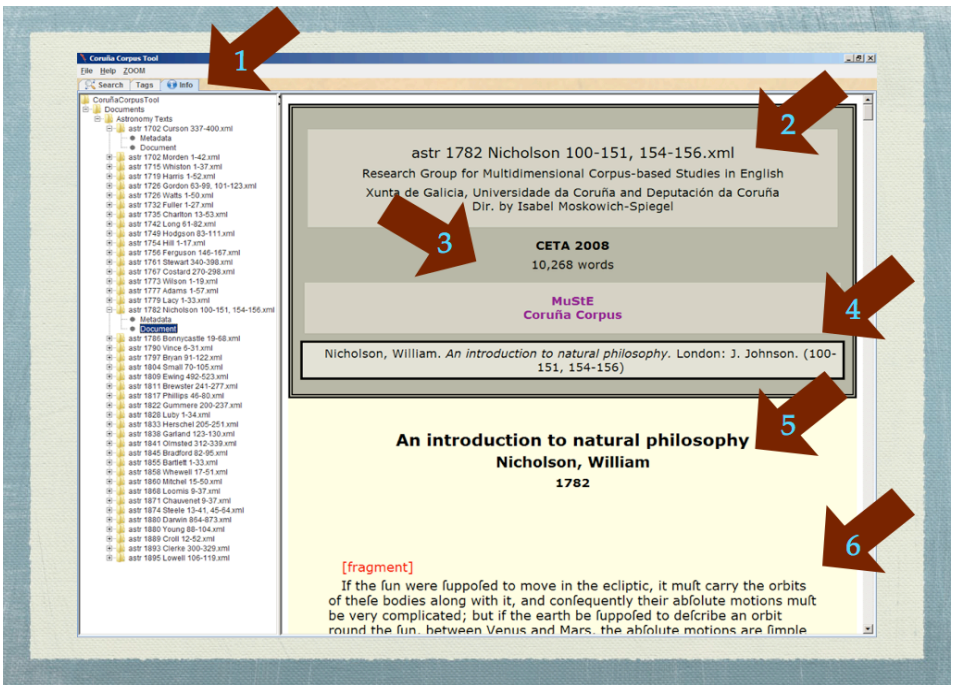


Figure 20: *Sample text view on CCT.*

Arrow three, however, gives specific information about this text, namely the subcorpus within *CC* this sample text belongs to (*CETA*) and the number of words the

text contains (10,268). Arrow four marks the bibliographical reference and arrow six, the text itself. All these features are a way of finding a standardized way of approaching the texts and help facilitating their reading. The study of particular linguistic features, however, remains quite arduous. For this reason, *CC* has been designed with the *CCT* search engine.

3.1.3. *The Coruña Corpus Tool (CCT)*

The *CCT* is a linguistic product developed by the IRLab (University of A Coruña) in collaboration with Muste (Lareo, 2009a, 2009b 2012; Parapar & Moskowich, 2007). The *CCT* was designed for joint use with *CC* and its main function is that of linguistic corpus management. The tool serves as a valuable validator for TEI encoded documents with support for non-standard characters. It supplies information about the format errors to allow the correction by the linguists. Apart from storing and validating text information, *CCT* is especially useful in searching for information. The *CCT* can generate concordances of simple words, phrases and sentences and locate them in the documents. Searches can be carried out within the whole corpus or within a specific subset that can be customized by the user. There is also the possibility of introducing a set of wildcards to search prefixes and suffixes and it is also possible to search for specific tags within texts, such as abbreviates and in-text quotations. The results of the searches are visualized within the tool and it is possible to generate types and tokens lists.

Figure 21: *The Coruña Corpus Tool.*

Given the nature of this study, a closed set of suffixes was selected to generate concordances. Luckily, the *CCT* includes a wildcard for searching suffixes: dot (.) asterisk (*) suffix. In figure 22, it can be seen how the input of the wildcard and one of the suffixes (.*tion) generated a list of concordances, ordered by text. Apart from the concordance list, the interface also shows the left and right context of the word, which is useful for disambiguation.

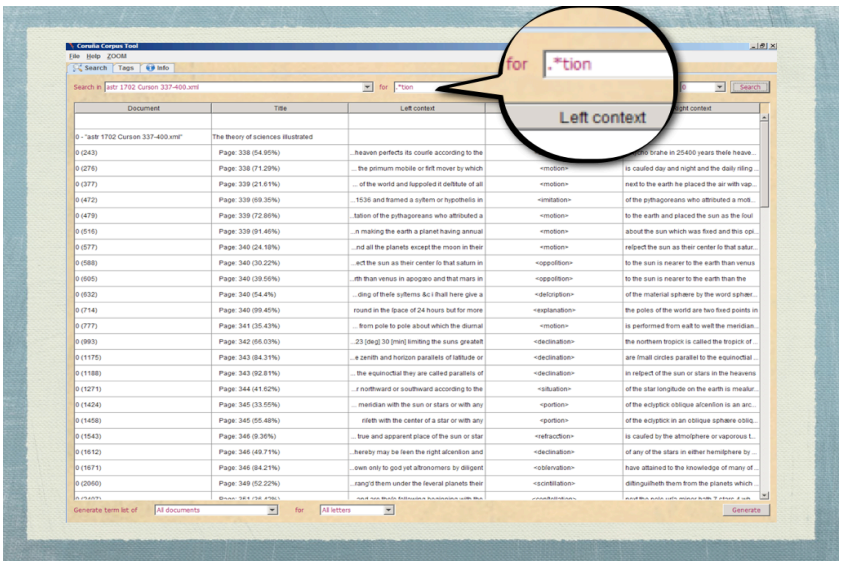


Figure (22). *Search in CCT.*

In the case that more information is needed about the context of the word, the exact location of the word can be accessed by clicking on the word, as can be seen in figure 22. Obviously, since *CC* is not lemmatized, the tool does not recognize the string /ing/ as a deverbal suffix and, consequently, the concordances generated would include gerunds, nominalizations or other words ending with these letters such as, for instance, *spring*. Consequently, and apart from the opportunities provided by the *CCT*, an extensive disambiguation process had to be later applied, as will be explained in section 3.2.

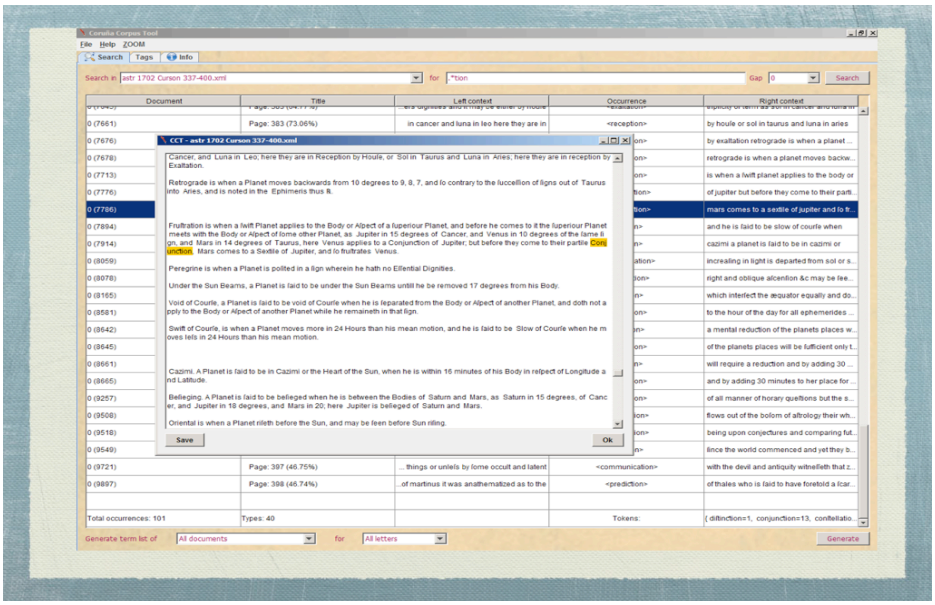


Figure 23: *Location of concordances in the text.*

By clicking on the last row at the end of the concordance list, the user can access a summary of the results, which is presented in a separate window. All the examples matching our query are presented in this window, together with the statistical data and the location within texts (page number, author and title of the text).

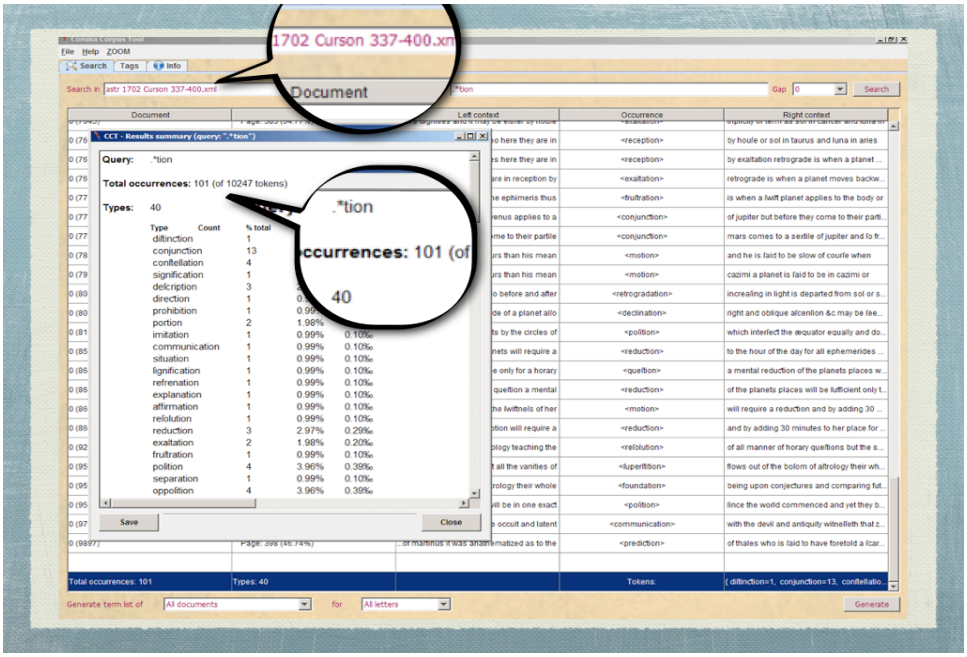


Figure 24: *Exploring CCT.*

In figure 24, the result summary window is shown. In this particular case, the query ".*tion" was introduced and the subset of texts was restrained to Curson's *Theory of Sciences Illustrated* from 1702 (see search box on the upper left corner). For this particular subset of texts, the *CCT* generated a list of concordances containing 101 tokens and 40 different types, which were classified alphabetically in a list and then located in their texts. The advantages of the *CCT* are obvious and it constitutes a real help for this study. Once the working tools have been carefully described, the next section will deal with the methodology used for carrying out this study.

3.2. Methodology

The process of corpus exploitation and database creation was indeed one of the most time-consuming tasks in this study. The whole process was staggered and designed to be fulfilled in progressive stages that served as advancement in the study and revision of the previous stages simultaneously. In this sense, the design of the study was not fully developed from the beginning but rather sketched, redesigned and adjusted as the study continued. It consisted of three stages: the first and the second were focused on disambiguation tasks whereas the third one was concerned with the creation of two specialized files, one containing data for extralinguistic analysis and the other, data for the study of nominalization typology based on morphological and syntactic variables.

Following the recommendations made by scholars about hypothesis testing and corpus linguistics (Taavitsainen, 2005) one of the main aims of this study was that of verifying the hypothesis that nominalization was an increasing linguistic feature in scientific register from the last 400 years. However, as the study advanced, new hypotheses were formulated during the process of corpus exploitation and database creation, which were later studied and whose results are presented in chapter four.

To avoid misrepresentation and to warrant some degree of homogeneity, in this first round all ambiguous cases were marked and saved for in-depth analysis. The objective of this first round was twofold: on the one hand, the tools available for this analysis did not refine the search to the point that human disambiguation could be completely excluded. On the other hand, manual disambiguation was time-consuming but extremely helpful for the establishment of the categorization of nominalizations that

was later used in the study. Figure 25 shows one of the fourteen sheets used in this first stage.

In the case of the examples contained in figure 25 out of thirteen tokens, nine (in red) were discarded automatically because they were either verbs (*presage*) or nouns not derived from verbs (*language, advantage, page, age*). If the corpus was tagged, the concordances of verbs with an ending coincident with the suffix used for this study would not appear in the result list. A disambiguation of nouns, however, would still be required to eliminate those nouns without semantic component of 'process'.

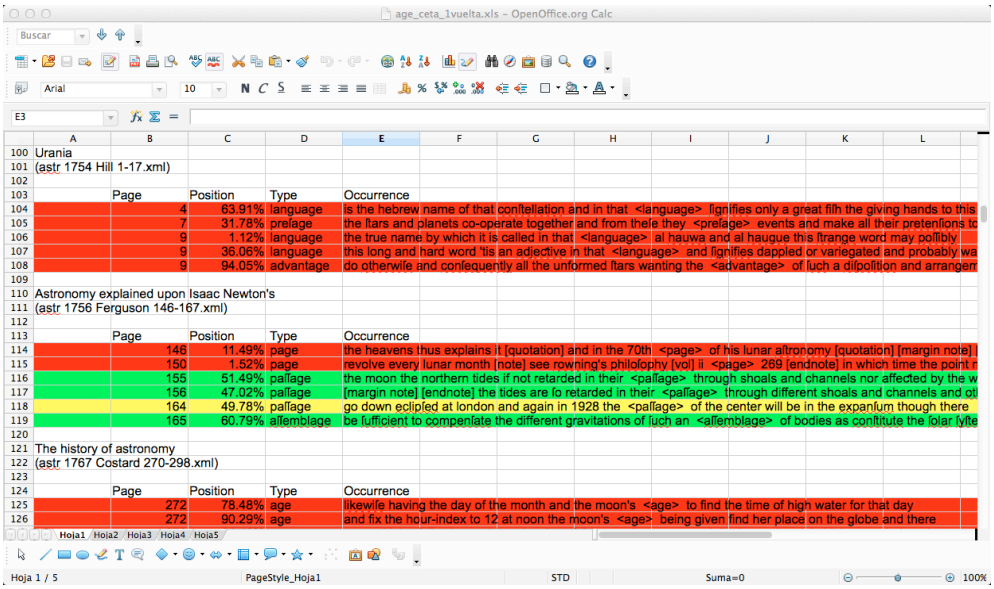


Figure 25: *First stage of disambiguation.*

Those nominalizations that could easily pass the morphologic and semantic requirements (marked in green) were validated whereas ambiguous cases (in yellow) were saved for later analysis. In the case of the example marked as ambiguous in figure 25, disambiguation could be done by widening and reading the context in which the word was inserted. In the second stage of the disambiguation process, the ambiguous

cases were carefully studied and either validated or rejected for analysis. The criteria followed for noun disambiguation responded to the following fields:

1. **Morphology (word class)**: only lexical nominalizations are covered by this study. The existence of determiners, premodifying APs and postmodifying PPs was actively used as disambiguators. Lack of modifiers, on the other hand, served as a powerful criterion for assigning them the "ambiguous" label on the first round, especially in the case of *-ing* nominalizations. Among examples (41) to (44), only the first one was considered a valid nominalization. The existence of a premodifying article and a postmodifying assignation clearly indicates that *passing* in (41) is a noun:

> (41) […] taken up by the **passing** of the spot many of these spots are very dark (Phillips, 1817, p. 47; emphasis added).

In (42), however, there are serious doubts for ascribing a category to "passing", as in this case the word has a premodifying NP and a postmodifying NP functioning as a direct object.

> (42) The act of a body's **passing** the meridian is called its culmination (Bartlett 1855, p. 19; emphasis added).

This merging of nominal and verbal features is very frequent in *–ing* nominalizations but out of the scope of this study. The doubts about word class are not so strong in (43) and (44) because it can be clearly observed that the NP and the PP in (43)

> (43) The rational horizon is the circle in which a plane **passing** through the earth's centre and parallel to the sensible horizon (Loomis, 1868, p. 12; emphasis added).

function as subject and prepositional complement of the verb *to pass*, which has been conjugated in gerund, whereas the context of *passing* in (44)

> (44) […] by the inflexion of the passing rays surrounded the shadows of all opaque globular bodies […] (Brewster,1811, p. 274; emphasis added).

indicates an adjectival function premodifying the noun *rays*.

2. **Morphology (word building)**: this study focuses on nominalizations formed by deverbal suffixation. There are two implications to this. First, nominalizations formed through conversion or those without a cognate verb will not be taken into account. This prerogative does not affect disambiguation, as the queries introduced in the search engine are actually the suffixes which these words lack. The second implication is that apart from deverbal nominalizations, other words with the same endings may appear in the list. In some cases, these endings would be only a matter

of coincidence and in others, the suffix may not be functioning as a deverbal suffix for the formation of a noun. A valid nominalization can be seen in (45):

> (45) [...] the equinoctial and its alternate **arrival** in the northern and southern [...] (Herschel, 1833, p. 205; emphasis added).

The noun arrival is formed of the root "arrive", which comes from Old French *ariver* (and this, from late Latin *arribāre*< *arrīpāre, ad to* + *rīpa* "shore" (*OED*) and the deverbal suffix *-al*. The word retains the semantic traces of a process, as can be perceived in the definition "the act of coming to the end of a journey, to a destination, or to some definite place" (*OED*). On the other hand, (46)

> (46) [...] you do me and my trifles a great **deal** of honour [...] (Harris, 1719, p. 22; emphasis added).

fulfills the requirement of word class but the ending *-al* cannot be considered a suffix because the word derives from OE *dǽl* and Old Germanic **daili-z*, "part, share, quantity, amount" (*OED*). In the case of (47)

> (47) [...] explain the real nature and cauſes of the **celeſtial** lights [...] (Charlton, 1735, p. 13; emphasis added).

the ending is indeed a suffix. The word celestial is a combination of this suffix and the root *celest-*, which derives from Latin *caelestis sky*. However, in this case *-al* is

functioning as a denominal suffix in the formation of an adjective, which places the word out of the interest range of this study.

3. **Semantics**: in this study nominalizations are understood linguistic expressions of a conceptual representation of a process or state of affairs in a nominal form, so nouns with a deverbal suffix lacking the semantics of 'process' were discarded. This was one of the more tricky stages in the process of disambiguation. Consultation with the *OED* and careful reading of the context in which the word appeared were the main techniques applied. Deverbal nouns that have been lost their verbal meaning are not under the scope of this study. Thus, nouns that are now listed in dictionaries under a different entries were, consequently, not taken into account. *Meaning* and *institution* are two clear examples of nouns that look like nominalizations but have undergone semantic shift. In the case of *institution*, the word appears five times in the corpus. Of those five times, four occurrences refer to an "established organization" and only one case corresponds to the action of instituting something. One example of *institution* with semantic shift may be (48):

> (48) Its formal opening and dedication to its scientific work formed the most interesting event of the meeting of the International Astronomical Society which occurred last summer at Berlin. This **institution**, provided with three large telescopes, [...] (Young, 1880, p. 104; emphasis added).

On the other hand, an example of a true nominalization keeping the "process" component can be:

(49) [...] the Gregorian Account has accordingly got before the Julian one Day more than it was at the Time of its **Inſtitution**, [...] (Fuller, 1732, p. 19; emphasis added).

In the case of "meaning", none of the forty-seven occurrences of the word correspond to a true-nominalization reading. Unlike *-ing* forms, in which it was easier to rely on formal features for disambiguation, determining the degree of verbal meaning kept in a noun is extremely fuzzy.

4. **Textual considerations**: in case of ambiguity, the context where the word is inserted is extremely important, and information about the main verb in the sentence can cast some light. In most cases a semantically-emptied verb is a clear indicator of the existence of a nominalization.

The percentage of ambiguous cases was 12.72% of the 25,771 words analyzed in the disambiguation process. After suffix breakdown, the highest percentages of suffixes with words discarded from analysis are those of *-al* (99,42%), *-ure* (92,94%) and *-ing* (91,81%). The most repeated reason for rejection was the coincidence with words with a different grammatical category. Thus, the suffix *-al* is more productive in the formation of adjectives, as in (50)

(50) […] that thoſe 5 hours and 49 minutes which the sun's annual **revolution** requires above 365 days will in 4 years' time […] (Watts, 1725, p. 11; emphasis added).

whereas *-ing* is typically found in verbs, as in (51):

(51) […] Perrotin was on the point of **abandoning** the search altogether when […] (Lowell, 1895, p. 109; emphasis added).

In the case of *-ure*, the most repeated case was *figure*, which appeared 309 times (almost the 25% of all the words ending with *-ure* in the corpus)[80].

(52) […] the ſame **figure** ſhews the ſtar s is 45 below the horizon […] (Long, 1742, p. 6; emphasis added).

The only suffix with a higher percentage of positive cases is *-ion*, with over 79% of positive cases among all the words with this ending in *CETA*.

Once the disambiguation processed was completed, a database with all 8446 nominalizations and all the variables of study, both linguistic and extralinguistic, was created. This database contains 8446 rows and twenty-three columns with all the variables, which makes up a total of 194,258 cells with information. Linguistic variables (columns K to W) were additionally numerically codified according to the closed set of possibilities available in each of them. Figure 26 contains a screenshot of this database:

[80] This can be explained by considering that figures are frequent elements of scientific texts.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
7056	1860	Mitchel, Ormsb	male	Ohio	academic treatise	reasoning	fade away in total ec	ing	Angl o-	1400	0	0	0	0	0	0	0	0	0	6	0	2	4
7057	1860	Mitchel, Ormsb	male	Ohio	academic treatise	refinement	earth's diameter and	ion	fren ch	1398	4	13	0	3	3	1	0	2	2	2	1	1	4
7058	1860	Mitchel, Ormsb	male	Ohio	academic treatise	revolution	the center of this circ	ion	fren ch	1393	0	0	0	0	0	0	0	0	0	6	0	1	4
7059	1860	Mitchel, Ormsb	male	Ohio	academic treatise	revolution	of this circle by watc	ion	fren ch	1393	0	0	0	0	0	0	0	0	0	6	0	1	4
7060	1860	Mitchel, Ormsb	male	Ohio	academic treatise	revolution	mercury by giving to	ion	fren ch	1393	0	0	0	3	4	0	1	0	0	6	0	1	4
7061	1860	Mitchel, Ormsb	male	Ohio	academic treatise	revolution	of the sun's apparen	ion	fren ch	1393	4	13	0	3	3	1	0	2	2	1	1	1	4
7062	1860	Mitchel, Ormsb	male	Ohio	academic treatise	revolution	moon's path crossec	ion	fren ch	1393	4	13	0	3	3	1	0	2	0	6	0	1	4
7063	1860	Mitchel, Ormsb	male	Ohio	academic treatise	revolutions	its effects by taking i	ion	fren ch	1393	4	7	0	0	0	0	0	0	0	6	0	1	4
7064	1860	Mitchel, Ormsb	male	Ohio	academic treatise	rotation	can only reach an ap	ion	latin	1500	0	0	0	0	0	0	0	0	0	6	0	1	4
7065	1860	Mitchel, Ormsb	male	Ohio	academic treatise	rotation	of venus the sun's ar	ion	latin	1500	4	2	3	0	0	1	0	1	0	9	0	1	4
7066	1860	Mitchel, Ormsb	male	Ohio	academic treatise	rotation	same distinguished (	ion	latin	1500	5	12	0	3	3	1	1	2	0	7	0	1	4
7067	1860	Mitchel, Ormsb	male	Ohio	academic treatise	solution	common origin here	ion	fren ch	1382	0	0	0	0	0	0	0	0	0	6	0	1	4
7068	1860	Mitchel, Ormsb	male	Ohio	academic treatise	solution	an effort to explain th	ion	fren ch	1382	4	1	0	0	0	0	0	0	2	2	1	1	4
7069	1860	Mitchel, Ormsb	male	Ohio	academic treatise	speculation	a central axis this im	ion	latin	1471	0	0	0	0	0	0	0	0	0	7	0	2	4
7070	1860	Mitchel, Ormsb	male	Ohio	academic treatise	speculation	and real magnitude l	ion	latin	1471	4	7	0	0	0	0	0	0	0	9	0	2	4
7071	1860	Mitchel, Ormsb	male	Ohio	academic treatise	statement	hundred or five hunc	me nt	fren ch	1789	4	1	0	5	0	0	0	0	1	2	1	4	4
7072	1860	Mitchel, Ormsb	male	Ohio	academic treatise	succession	own earth these will	ion	fren ch	1405	1	0	0	0	0	0	1	0	0	6	0	1	4

Figure 26: *The database.*

All the variables contained in this database were later used as variables in the study. They included:

1. **Chronology**: this parameter responds to one of the central aims of the whole study, that is, the verification of the claim that the use of nominalizations in English scientific register in the last three centuries has steadily augmented and become of its distinguishing features today.
2. **Gender of author**: this field can only be filled with the tags "male" or "female." One of the central focuses of interest in this study, the ultimate goal intended with this parameter is to delimit whether and to which point language use is different according to sex. Expectations about results in this field are high, as the general hypothesis assumed is that the description of scientific English register has been modeled according to male writers, who were the main, almost exclusive, writers of science three centuries ago.

3. **Author's place of education**: the cutting point in this parameter has been the country in which authors acquired their linguistic abilities. The aim of this parameter is to determine to which point the use of nominalizations in scientific writing differs in different regions. The hypothesis is that the number of nominalizations should increase at a higher pace in Europe (first in England) in the eighteenth century because this was the place where the Scientific Revolution originated.
4. **Text-type**: the labels used for this field coincide with the ones proposed by *CETA* compilers. Nominalizations are considered a feature of a particular register. The underlying hypothesis that serves as a starting point is that there will be more nominalizations in the most learned text-types because of their value as scientific discourse markers that members of the discourse community are expected to include. Besides, due to valency reduction and condensation, nominalizations are more difficult to decode and need more trained readers.
5. **Suffix**: the seven suffixes included in this study, namely *-age, -al, -ance, -ment, -ing, -ion*, and *-ure*, are taken as variables as well. The main objective of this variable is to determine what is the most productive suffix for the creation of deverbal nominalizations.
6. **Etymology (origin and year of introduction)**: for these variables, consultation of the *OED* was carried out. The etymological variable will cast some light on whether the Germanic or the Romance element has been more productive in the creation of nominalizations and to which point blending of roots and suffixes with different origin was a common word-formation process. On the other hand, the variable of "year of introduction" will help determine the degree of innovation in the increase

of nominalizations in scientific register: were nominalizations already a standardized linguistic feature and the only innovation was its increase in frequency in scientific register or was there a drastic change and the development of science was correlated by the springing of a new word-formation process?

7. **Premodifier**: the first parameter about the structure of NPs governed by nominalizations is the type of determiner. A numeric code was given according to the eight available options for premodification. There are no hypothesis for this or for any of the morphological variables, as their only aim is to fid out what is the most repeated value in each category and if there are structural restrictions within nominalization NPs.
8. **Type of determiner**: since determiners were the most common type of predeterminer, this variable aims at establishing what types of determiners are most frequently combined with nominalizations. The list of determiners was based on the typology established by Quirk et al. (1985).
9. **Possessive structure**: possessives are, after articles, the second most productive determiner premodifying nominalizations. They are indeed a great source of information of pragmatic and functional features that can help divisions among the different typologies. Possessive constructions often add complexity and ambiguity and their combination with nominalizations normally results in intricate phrases.
10. **Postmodifier**: following the structural analysis of the NP, this variable analyzes the type of postmodification usually found with nominalizations. Broadly speaking postmodification can be lexical or clausal and it is expected to find higher frequency rates for lexical postmodification.

11. **Function of postmodifying PP**: PPs are the most common type of postmodification. In this variable all its possible functions are numerically codified. The type of function is delimited by the hypothetical function that PP would most likely fulfill in case of a verbal realization.
12. **Valencies in common with VP**: in this variable, the number of modifiers is put in relation with the semantic role they have in the phrase (agent, participant, object, circumstance) and with the relationship they maintain with the verbal valencies of an equivalent verbal realization.
13. **Circumstance inclusion**: one of the main claims of this study is that nominalizations are not verbal transformations, but rather respond to a different mental conceptualization of a process. This may be proved by circumstance inclusion. In the VP, adjuncts are optional elements so they are less likely to be transformed into nominal modifiers. Consequently, my expectation is to find a good number of circumstance inclusion to prove that VP structure does not shape the structure of NPs governed by nominalizations
14. **Agency inclusion**: the expectation is that agency will be the most common element included in the modifying field. In this variable, not only the presence or absence of an agent, but also its position in the phrase is taken into account.
15. **Type of clause**: whether the clause in which the nominalization appears is subordinated is taken as a typology indicator. Thus, it is expected that more thematic nominalizations will appear in main clauses, whereas conditioned nominalizations will most likely be found in subordinated sentences. The category "non-applicable" was also introduced for those cases in which the nominalization functions as a modifier within another element.

16. **Syntactic function**: all the possible syntactic functions are considered in this variable and the expectation is that nominalizations will most likely function as subjects or subject complements, following Halliday's (2004) teachings.
17. **Existence of semantically-emptied verb**: another feature of scientific register mentioned by Halliday (2004), the existence of a semantically-emptied verb may indicate either a light-verb construction (stylistic nominalization) or a thematic nominalization functioning as subject. This parameter has been included because it can help disambiguate different typologies although it will not be shown in data analysis.
18. **Typology**: after the reading of all linguistic features as well as the context in which the word is inserted, each nominalization was assigned a typology. This variable then served as the common ground for the second part of the analysis, where contrastive typological analyses considering most of the morphosyntactic variables previously described were also carried out.

In this chapter, the corpus and methodology used for data analysis has been presented. It has been shown that *CETA* is perfectly appropriate for the type of analysis intended up to the point that some of the study variables, such as text categorization and place of education, have been modeled upon the sociolinguistic information contained in the metadata files. However, it has also been made clear that not all the compilers' choices have been followed. Consequently, in the geographical variable, analysis will primarily focus on continental variation and then it will be segmented in three groups (England, Ireland and Scotland) for European and four groups for American (New England, Midwest, Mid-Atlantic and South Atlantic). Concerning the methodology

followed, the two main steps, disambiguation and database creation, have been explained. The results of the latter will constitute the heart of the core of next chapter.

4. Analysis of data

As advanced in the title, this chapter constitutes the analysis of the nominalizations found after disambiguation processes and database creation. Several variables have been applied. First, the total number of nominalizations were subjected to extralinguistic variables (section 4.1.1). The extralinguistic variables included have been: chronology (section 4.1.1.1), sex of author (section 4.1.1.2), place of education (section 4.1.1.3) and text-type (section 4.1.1.4). The following section (4.1.2) consists in the application of linguistic variables. This analysis is faced at two different levels: whereas section 4.1.2.1 is concerned with the morphosyntax of nominalizations, section 4.1.2.2 deals with the morphosyntax of the NPs in which those nominalizations are inserted.Once all nominalizations have been analyzed as a whole, an analysis according to the typologies presented in section 2.5 is presented in section 4.2. The typological analysis is slightly shorter and some variables of study have been left out of the present work. Only those variables that presented some significant variation have been included with the aim of maximizing conciseness. The structure of the linguistic analysis, however, remains

identical: after an analysis of the morphosyntax of the different typologies (4.2.1), the emphasis is drawn to the structure and functions of nominalization NPs (section 4.2.2).

4.1. Analysis of nominalizations

After corpus exploitation the total number of nominalizations considered for analysis has been 8,446, which represents 4.17% of the 20,2403 words contained in *CETA*[81]. In this chapter, the analysis of those 8,446 nominalizations is carried out following two axis. First, a set of extralinguistic and linguistic variables are applied to the total number of nominalizations (section 4.1). Then, the same variables are applied to the four typologies described in chapter two (section 4.2).

4.1.1. Extralinguistic variables

Extralinguistic variables study how sociological factors can cause linguistic variation. Through them, the relationship between language and society can be placed in its historical dimension. One of the main aims of this study is to try to account for historical change concerning the use of nominalizations in scientific register in late Modern English and sociolinguistic variables suit this end. *CETA* contains, along with

[81] Unfortunately, at the time this study was made, *CETA* was not parsed, so the percentage of nominalizations with regard to the total number of nouns cannot be provided.

the sample texts, a series of "metadata files" that allow sociolinguistic analysis. The extralinguistic variables selected for this study have been chronology of texts (section 4.1.1.1), sex of author (section 4.1.1.2), place of education of authors (section 4.1.1.3) and text-type (section 4.1.1.4).

4.1.1.1. Chronology

Given that this is a diachronic study of scientific English, the first parameter to be analyzed in this section is chronology. Figure 27 offers a visual representation of data concerning the number of nominalizations found in texts across the two centuries under study.

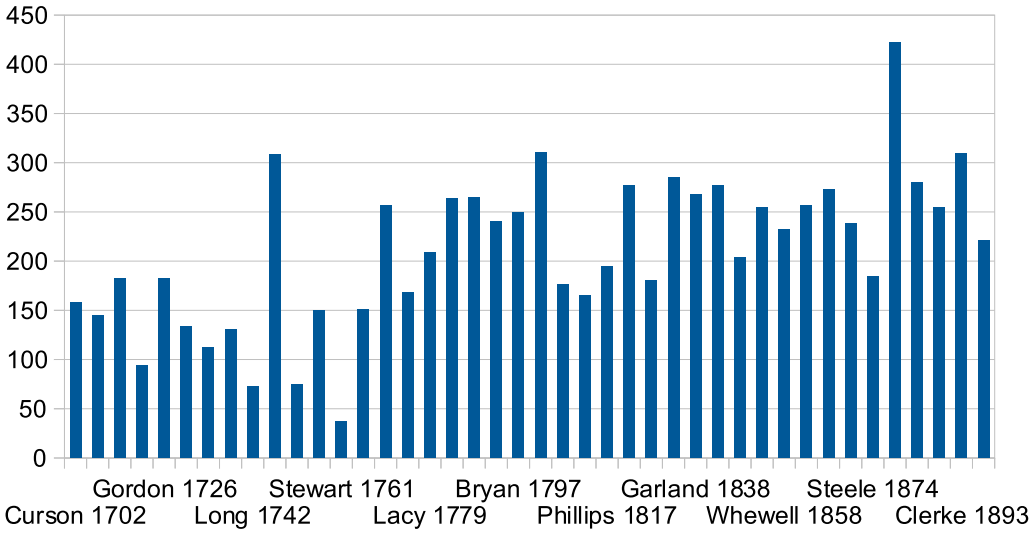


Figure 27: *Evolution in the frequency of nominalizations.*

It can be easily remarked that the tendency is that of a slight increase in the number of nominalizations. There are however some atypical results which do not interfere with the general trend of steady increase. In the fist half of the eighteenth century, the mean number of nominalizations per text is 151. Long's textbook and Hill's dictionary, with a mean frequency of 76 and 75 respectively, are slightly under that number but, as will be seen later, these two are learner's text-types and tend to include less nominalizations to avoid grammatical and semantic complexity. The case of Hodgson's textbook from 1749 is, on the other side, highly atypical, as with a mean frequency of 342 nominalizations, greatly outnumbers the usual number for its period. The opposite situation can be found in the second half of the century, more specifically in Stewart's essay from 1761. The mean frequency for this period being 181 tokens per text, the 45 nominalizations found in Stewart´s seem rather atypical, especially if we consider that this is a formal text, whose frequency expectation is higher[82]. The rest of the period looks under the norm, with the mean frequency raising over 250 in the twenty last years of the century.

The mean frequency for the first half of the nineteenth century is 234 tokens per sample text. Excepting the 324 nominalizations found in Small's academic treatise from 1804 and the 162 tokens from Brewster's treatise from 1811, all texts do not deviate excessively from the mean frequency. The last half of the nineteenth century presents the smallest growth in mean frequencies, with 239 tokens per sample text. Similar to what happened in the first half of the century there are not big oscillations and the lowest peak –Steele's textbook from 1874– differs only in 55 tokens from the mean frequency for the period, which cannot really be considered a major discrepancy.

[82] More information about text-types and nominalization frequency is to be found in section 4.1.1.4.

The total amount of occurrences is shown in figure 28, grouped in periods of twenty years[83] –each period contains four texts. In this graph, the progressive increase in frequency and the quantitative change from the last twenty years of the eighteenth century until the end are even more remarkable.

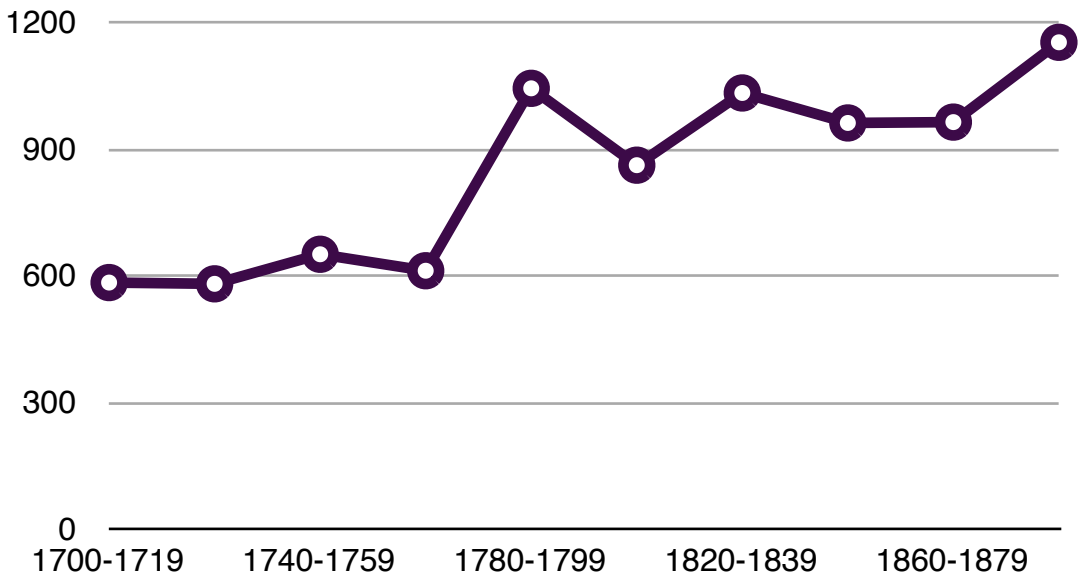


Figure 28: *Evolution in the frequency of nominalizations (twenty-year groups).*

Concerning possible motivations for this progressive augmentation, there is no apparent reason to believe that this change was motivated by the subject of study itself. Scholars acknowledge that nominalizations are complex linguistic devices (Halliday, 1985, 2004; Ventola,1996). Hence it would be reasonable to expect that the increase in the complexity of the scientific topics discussed in texts is linked to the increase in the use of complex linguistic structures. However, it would be very risky to claim that the use of nominalizations is linked to the complexity of the topics of the texts and that

[83] Even if authors such as Siemund and Claridge (1997) indicate that it is advisable to study language change in periods of thirty years, the choice of five groups of twenty years each per century instead of groupings of thirty years has been preferred to facilitate study across centuries and the establishment of five groups per century.

texts from the nineteenth century are more complex than those from the eighteenth century. Examples (53) and (54) may be said to be similar in form –with minor spelling differences–, function and meaning, although they are from different centuries:

> (53) Let me warm you a little with this **Deſcription** of theſe Zones given by Mr. Dryden […] (Harris, 1719, p. 44; emphasis added).

> (54) But before proceeding further with a **description** of these Martian phenomena, the history of their discovery deserves to be sketched […] (Lowell, 1895, p. 109; emphasis added).

In both cases, the nominalization appears in a PP functioning as adjunct and is premodified by a single determiner. Additionally, both examples have a postmodifying PP that contains the object of the process expressed in the nominalization.

There is further evidence showing that the claim that an increase in the use of nominalizations is not necessarily motivated by an increase in the complexity of texts. As already mentioned in section 1.2, the development of astronomy was slower in America than it was in England, mainly as the result of all the social and political struggles that hindered the development of scientific disciplines in the U.S. Consequently, if we consider that the number of texts written by American authors in *CETA* is greater in the nineteenth century and that there is no reason to believe that American texts are more complex, but rather the contrary, there is no reason to believe that the use of nominalizations is linked to the complexity of the text itself.

The motivation for this steady increase in frequency must, therefore, come from another side like, for instance, register and discourse community. For reasons alien to the text itself, nominalizations started to be used as markers of scientific discourse. At

the same time, they can also be regarded as stylistic devices carrying some a certain degree of guild codification. The inclusion of nominalizations in texts would make writers feel they belong to a discourse community because they master not only the message but also the code they are using to communicate it. Thus, it may be argued that the progressive augmentation in the use of nominalizations shown by data is the result of the evolution in the existence of specialized registers in English. The acquisition of a progressively codified register could be linked to the effects of the Scientific Revolution and the evolution of science on a general scale. Thus, we find that movements of institutionalization of science that took place in the seventeenth century were paralleled to the development of a scientific register displaying features of its own.

Connected with the evolution of nominalizations is the relationship between types and tokens. Even if some scholars (Lyons, 1977) have claimed that linguistics should only be concerned with types, that is, word-forms, corpus linguistics also pays a great deal of attention to tokens, that is, the occurrences of those word-forms. In this study, the number of nominalizations are repeated in the corpus is as valuable source of information as the number of word-forms included. Indeed, studies (Bello, 2010b) have shown that the difference between text-types and tokens is relevant [84]. Figure 29 shows the total number of nominalizations in terms of types and tokens.

[84] The variable of type/token distinction can indeed be extremely significant. This was one of the main conclusions of a preliminary contrastive analysis I carried out in 2010 (Bello, 2010b). The application of this variable to a contrastive analysis of nominalizations in two *CC* subcorpora (*CEPHIT* and *CETA*) showed different trends in nominalization distribution. The total number of nominalizations was similar for both disciplines but, whereas philosophy texts featured more types, astronomy texts tended to have lower type and higher token frequencies. The sociolinguistic reading of this had to do with the education of authors and the topic of texts: philosophers received a broader education in classical languages and that had a reflection on their writing. On the other hand, astronomers, especially in the nineteenth century may not be fluent in so many languages and the use of nominalizations in their wirtings is more technical and has less type variation.

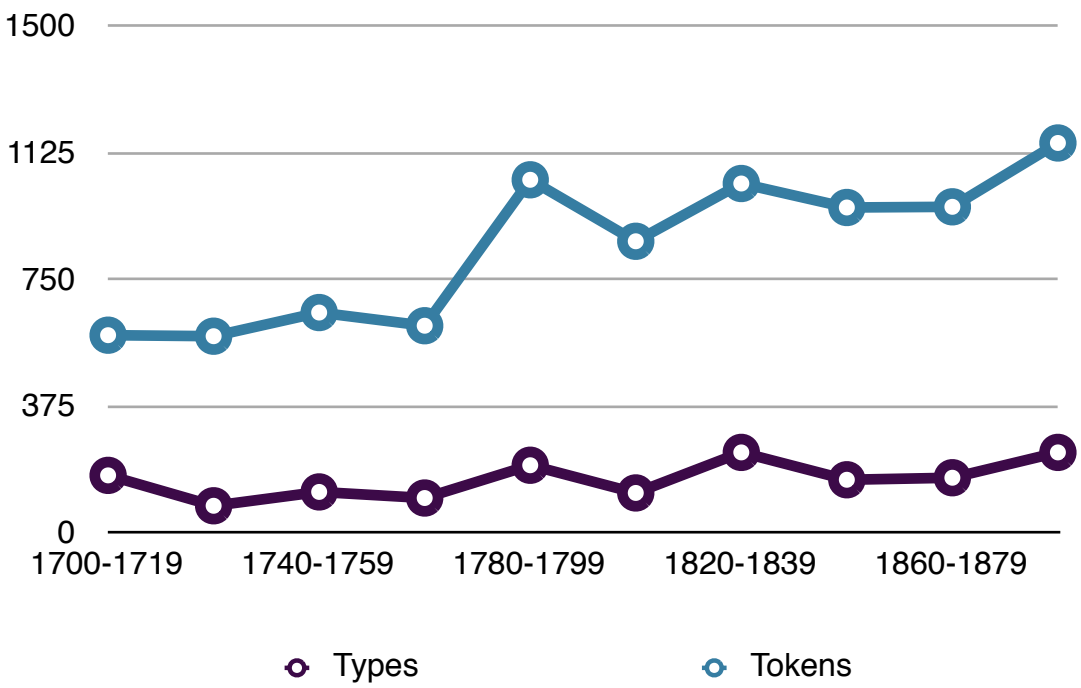


Figure 29: *Evolution of types and tokens.*

The data contained in figure 29 shows that the increase of nominalizations affected not only the total number of occurrences but also the number of repetitions in the same text. In the nineteenth century a wider variety of nominalizations was used in texts and these were repeated more frequently. This clearly indicates that nominalizations progressively became a prominent marker of astronomy discourse. Across the two centuries the number of types appearing only once (*hapax legomena*) is 227, and there is also a good number of types (14) which have more than 100 hundred appearances. In the first half of the eighteenth century nominalizations tended to appear very few times in texts –around one or two appearances per type–, towards the end of the nineteenth century, the trend changes to a mean of 15 appearances per sample text. The number of *hapax legomena* at the end of the nineteenth century is still high, which points out at the normalization of nominalizations as linguistic markers, as their increase

is not only in number but also in the ways they are used in texts. The two analyses presented in this section show that Halliday's (2004) claim that the tendency in scientific register in English was a steady augmentation of nominalizations is grounded. Additionally, the study of types and tokens has showed that nominalizations increased its number and established themselves, as discourse markers as the increase affects not only tokens but also types.

4.1.1.2. Sex of author

The gender variable may have the risk of being slanted, as there are only two texts signed by women in the corpus, As discussed in chapters one and three, this is a reflection of historical reality. Figure 30 shows the distribution of normalized frequencies of use of nominalizations in texts written by men and women respectively.

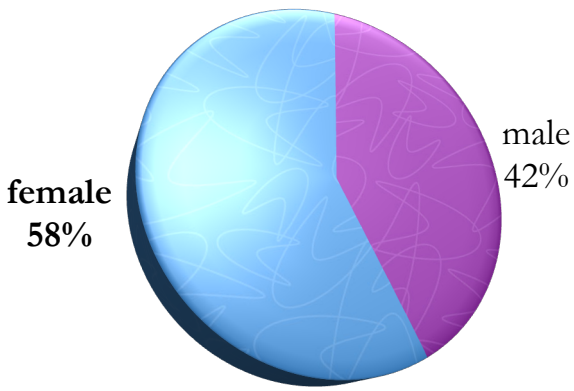


Figure 30: *Nominalizations according to sex of the author (NF 10,000).*

In this case, the balance slightly leans towards the female side, with a mean frequency of 279 nominalizations per sample text, while the mean frequency for male sample text remains 202. At the time when texts were written, women were believed to be incapable of showing abstraction on their thoughts and, consequently, on their writing. Their writings were not considered very informative and their style, loose (Moskowich & Monaco, 2014). Indeed, female writing was usually associated with a flourished style, more apt for literature and old approaches to science, whereas men's writing was though to be more concrete and abstract, which made it more appropriate for the new science Studies on women' writings in *CC* (Crespo, 2011; Moskowich, 2013; Moskowich & Monaco, 2014) showed that these hypotheses were mainly mistaken. Eighteenth-century women use an impersonal style that points out their ability to express abstraction. This feature is particularly evident in astronomy texts, which, due to their topic, are more prone to deal with abstraction than other Humanities disciplines. Moskowich (2013) showed that an abundant use of passives was a clear marker of an impersonal, abstract style. In this sense, nominalizations recall the same idea. They are complex discourse markers that normally draw attention on the process while attenuating the reader's attention on the agents, objects and circumstances related to that process. In many cases, nominalizations are also reified, which is also an unambiguous feature of abstract reasoning.

Women's tendency to include more nominalizations in their texts can thus be seen as a complex way of gaining weight in the male-centered academic world. Studies (Crespo, 2011, in press 2014a; in press 2014b; in press 2014c; Curzan, 2011) have shown that women writers did not resort to overt expressions of persuasion. Despite their unequal situation, their writings had low percentages of features denoting

involvement. After analyzing the use of predictive modals, conditional subordination and necessity modals it was discovered that men resorted to modality much more than women. So the high nominalization frequencies found in the corpus can be read as a veiled way of claiming credibility as skilled scholars: on the one hand, nominalizations added abstraction to texts. On the other, they were markers of scientific register. By mastering the language of science, they were claiming that they could also master research. It is not possible to delimit to what degree the decision of using nominalizations was a planned way of gaining credibility. However, it is always possible that, as avid readers of scientific texts, they probably came to master the flourishing register and reproduced it in their own writings. In any case, it is clear that the frequent use of nominalizations found in texts signed by women shows that in spite of their exclusion from official science, their style had assimilated the conventions associated to register. Nevertheless, given the small number of texts analyzed, the validity of this claim should be tested on a bigger corpus to assess that these results are not influenced by personal idiosyncrasies.

4.1.1.3. Place of education

In this section the place where authors acquired their linguistic skills is used to determine to what extent nominalization use is dictated by the geographical variable. The analysis of nominalizations according to the continent of origin of writers is summarized in figure 31. Some considerations must be made to help interpret the data

in this chart. First, American texts only include texts written by authors from the United States. No Canadian authors have been included in *CETA*. Similarly, under the European label there are only texts by English, Scottish and Irish authors. Other English-speaking nationalities –either European, American or from other continents– are not included in *CETA*, and, consequently, are not reflected in this graph.

The reason for creating two main groups according to continent instead of country of origin is not related to geographical reasons. Making distinctions between Scotland, England and Ireland on the sole basis that, unlike American states, they were independent nations would ignore the basics of dialectal variation. Even if standardization practices restrain dialectal variation, it is highly probable that the difference in language use between a pair of authors from Newcastle and Glasgow may be smaller than that of another pair of writers from Boston and San Francisco. Examples (55) and (56) have been taken from texts published in the same year in different continents. Example (55) was produced by George Darwin, son to the famous biologist Charles Darwin, born in Kent and educated at Cambridge (England). Example (56) was written by Charles Augustus Young, a university professor born and educated in Hanover, New Hampshire.

> (55) We are rather here merely concerned with those elements which contain a **description** of the nature of the orbit (Darwin, 1880, p. 864).
>
> (56) At any rate, it is evident that the subject requires **a re-examination** with instruments of higher dispersive power than any hitherto employed (Young, 1880, p.100; emphasis added).

(55) and (56) are similar not only in their structure but also in the function they play in the text: they are both stylistic nominalizations functioning as direct objects in a relative clause and they are postmodified by PPs expressing the object of the process in (55) (*the nature of the orbit*) and a circumstance in (56) (*instruments of higher dispersive power than any hitherto employed*). Their similarity shows to what degree differences according to geographical variation can be misleading. As a result, and given the fact that this piece of work does not aim at providing an exhaustive study of dialectal variation in the use of nominalizations, it seemed appropriate to study nominalizations according only to continental variation. Together, the results of the study according to place of origin can be seen in figure 31 below:

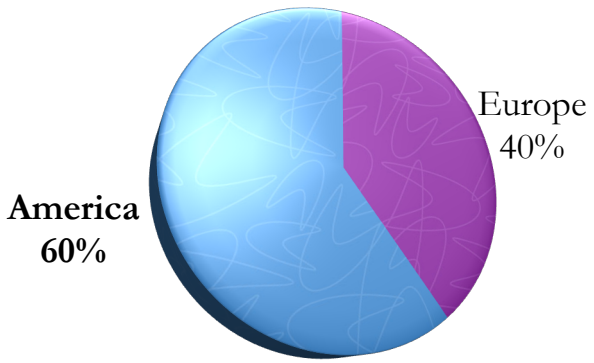


Figure 31: *Nominalizations according to author's place of education (continent) (NF 10,000).*

Texts produced by American writers contain more nominalizations. Whereas European texts have a normalized mean frequency of 175 items per sample text, American texts have a normalized mean frequency of 268 tokens. Looking at the results it might be asserted that American scientific register is more prone to use nominalizations as discourse and in-group markers. However, several factors may undermine the accuracy of this claim. First, it may be noted that no texts written in the

eighteenth century by American authors have been compiled in *CETA*[85], mainly due to extralinguistic reasons. The data contained in the figure reflects normalized mean frequencies, not absolute values but, as we have already seen, the number of nominalizations tended to increase in the nineteenth century. As a result, it is not clear whether the high frequency found in American texts is related to dialectal variation or to language change across the centuries.

It is true, however, that the rate of general evolution in the use of nominalizations shown in figures 27 and 28 is slightly smaller than the difference in the use of nominalizations according to continent shown in figure 31. This may lead to conclude that in spite of all the reasons previously cited, American texts favor slightly the use of nominalizations.[86]

Once continental variation has been analyzed, a closer look at the use of nominalizations in each separate continent will follow. Figure 32 shows that the percentages of normalized mean frequencies for European nominalizations is roughly even.

With a total number of 633 nominalizations and a mean frequency of 201 nominalizations per sample texts, the three Irish writers lead the group of most prolific European nominalization users. After them, Scottish authors use a mean of 183 nominalizations per 10,000 words and authors educated in England limit their use of nominalizations to 169 items per sample text.

[85] The first text by an American author is Ewing's, from 1809.

[86] It might be interesting to follow this thread of research, widen up the dates of analyzed texts and study to what degree continental variation is related to the abundance of nominalizations in standard scientific English register today.

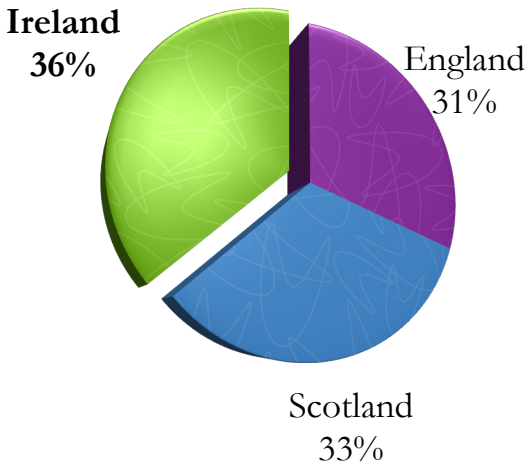


Figure 32: *Nominalizations according to author´s place of education (Europe only) (NF 10,000).*

The distribution of frequencies for American authors expressed in figure 33 seems equally egalitarian.

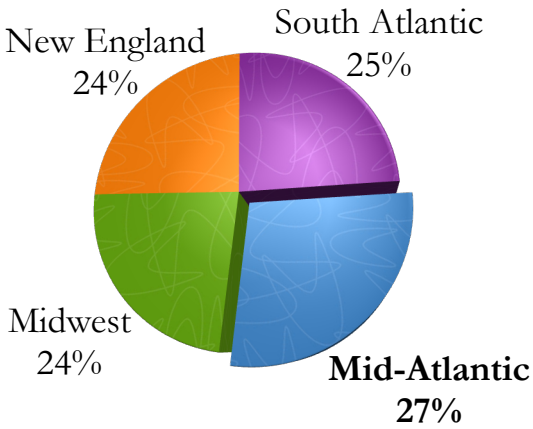


Figure 33: *Nominalizations according to author´s place of education (America only) (NF 10,000).*

In the case of American authors, the mean of nominalizations per sample text is highly superior[87], ranging from 257 nominalizations per sample text from the Midwest area to 296 tokens in Mid-Atlantic texts. Mid-Atlantic states (New York and Pennsylvania) overtake New England and Ohio (Midwest) by two points but trying to find a reasoning behind these data seems unwise. Looking at the small differences both across and within continents, the geographical variable does not seem very indicative in this study.

4.1.1.4. Text-Type

The study of language variation according to text typology is one of the pillars for the creation of the corpus. Compilers of the *CC* distinguish eight different text-types, namely academic treatise, dialogue, dictionary, essay, lecture, letter, research article and textbook. Figure 34 shows the distribution of mean frequencies according to text-type.

It can be noted that the text typologies with the highest number of nominalizations per sample text are research articles with 281 tokens, closely followed by letters with 270 tokens per sample text. After these, only the mean frequency of nominalizations found in academic treatises exceeds 200 tokens. On the other part of the scale, the only typologies computing less than 100 nominalizations are dialogues (93 tokens) and dictionaries (75 tokens).

[87] As explained before, this is a consequence of text selection, as all American texts in the corpus were written in the nineteenth century. The difference when compared to European mean frequencies is, therefore a result of chronological rather than geographical parameters.

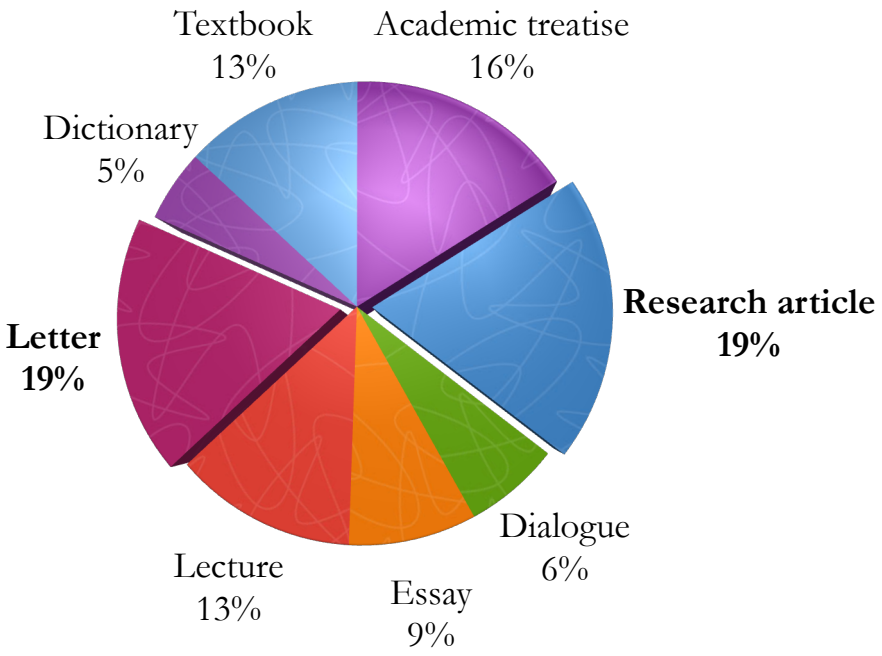


Figure 34: *Nominalizations according to text-type (NF 10,000).*

This distinction has been narrowed down to only two categories for the sake of simplicity. As in the case of geographical variation, this study does not aim at providing a comprehensive account of the evolution of nominalizations according to specific text-types. Consequently, from the initial eight text-types established in the *CC*, two categories encompassing formal and learner texts have been set up. Table (6) shows the text-types included in each category:

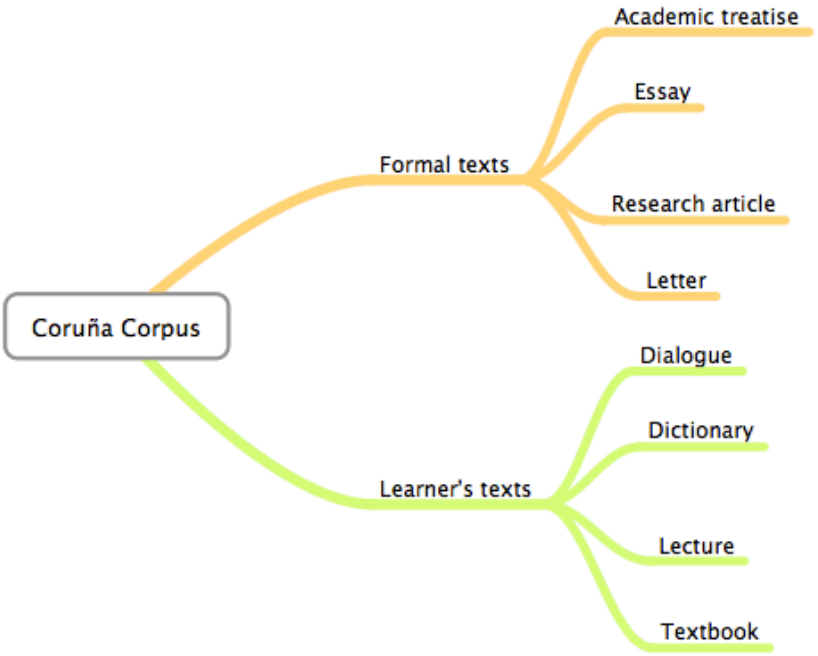


Table 6: *Formal and learner's texts.*

All the possible text-types were divided into two main groups. On one hand, formal texts are aimed at the scientific community. Language use in these texts is *a priori* more complex and sophisticated, whereas learner texts are targeted at a non-specialized audience, either students or the general public. Hence, in learner texts it is more probable to find not only less specialized vocabulary but also less complex linguistic structures. Data about the target audience may be found in the title of the works and in the prefaces written by the authors themselves.

Figure 35 shows the percentage of nominalizations found in formal and learner's texts. The number of nominalizations is considerably higher in formal texts –the normalized mean frequencies are 911 tokens per text versus 541 in learner's texts.

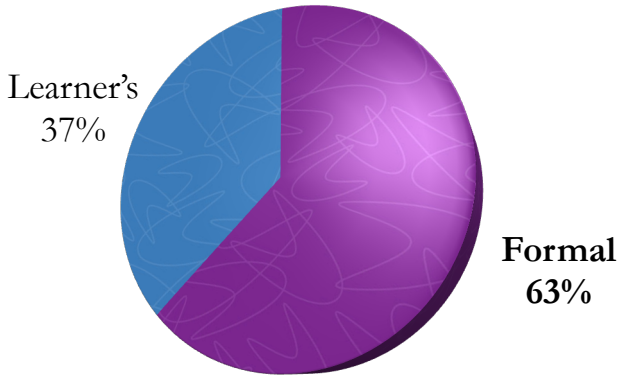


Figure 35: *Nominalizations according to intended addressee.*

This results confirm the expectation that the frequency of use of nominalizations in formal texts should be higher. If nominalizations are complex linguistic devices and they add complexity and ambiguity to texts (Banks, 2005a, 2005b; Guillén, 1998; Halliday 1985, 2004; Ravelli, 1988; Ventola, 1996), it is expectable that their frequency will be higher in texts aimed at a learned community that does not need an extremely simple style to understand what is being said. Furthermore, a complex style and a heavy use of specific discourse markers is also desired as a sign of belongingness to a discourse community.

Here again, a socio-external reading of the data can also be reached. The high number of nominalizations in formal texts can be said to be a consequence of the progressive dissemination of knowledge that was going on in those centuries. The augmentation in the number of scientific societies and groups of scholars called for the increase in nominalizations, which served as powerful in-group discourse markers for the creation of epistemic communities. This fact lays in close relation with the establishment of stylistic guidelines regarding specialized vocabulary, morphosyntax and text structure for the writing and publication of observational or experimental papers after the founding of the Royal Society (Crespo, 2011). Similarly the subsequent

institutionalization of science and the subsequent specialization of the sciences and of the language used to communicate them can also be considered a major motivation for the standardization of those stylistic guidelines. Grammar and stylistic complexity can be linked to scientific communities in the sense that these are reasonably closed sociolinguistic groups sharing particular linguistic uses. According to data in this study, nominalizations can be confirmed to be one of those stylistic markers of scientific English.

In this chapter it has been shown that the frequency of nominalizations, as expected, progressively increased. Also, as it was expected, the level of formality is correlated with nominalization frequency, which points out at the establishment of nominalizations as a type of guild codification that would differentiate learned and learning audiences. However, contrary to my expectations, it has been shown women tended to use more nominalizations than men. A temptative conclusion can be that this was a subtle strategy women used to reivindicate the validity of their writings. By using nominalizations they were making clear that they could achieve abstract thought and that they could also master the new style for writing science. Finally, the geographical variable shows that American authors were more prone to include nominalizations although this is rather a consequence of the fact that U.S. astronomy did not reach European standards until the nineteenth century and by that time, scientific register standards had already been established.

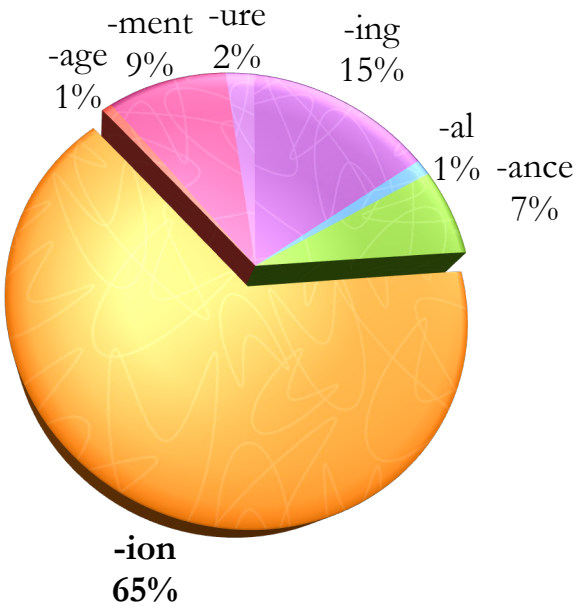
4.1.2. Linguistic variables: the morphosyntax of nominalizations

This section covers linguistic parameters of study of all the nominalizations extracted from the corpus. The variables of study include morphological and syntactic issues. Concerning morphology, the structure of nominalizations is studied in section 4.1.2.1. In it, parameters like suffix and etymology are studied. Then, in section 4.1.2.2 attention is drawn to the syntactic functions fulfilled by nominalizations and by NPs governed by nominalizations in sentences.

4.1.2.1. The morphology of nominalizations

As stated before, this study is concerned only with deverbal nominalizations formed through suffix addition so, analysis according to suffix is pertinent in order to have a general vision about nominalizations. Figure 36 shows the distribution of types according to the use of the seven suffixes that were chosen for this study.

It is remarkable to realize that five of the suffixes under study are scarcely productive. The suffix *-al* computes only 30 appearances that belong to eight different types, namely *arrival* (11 tokens), *disposal* (2), *perusal* (2), *removal* (2), *reversal* (1), *revival* (1), *trial* (10), *upheaval* (1).



Graph (36). *Nominalizations according to suffix (types).*

A little more productive are *-age*, with 47 tokens and five different types – *assemblage* (3), *passage* (24), *patronage* (1), *ravage* (1) and *voyage* (16)–, and *-ure*, which computes 89 tokens and a total of fourteen types. Suffixes *-ance* and *-ment* exceed the barrier of 300 appearances: there are 337 nominalizations formed with *-ance* belonging to 43 different types, while the 59 types of *-ment* nominalizations are spread in 464 appearances.

The suffix *-ion*[88], with 413 different types is clearly the most productive in the list[89] and were we reflecting the number of tokens with this suffix, its predominance would be even greater, as 82% of the occurrences of deverbal nominalizations formed through suffixation in *CETA,* that is 6936 tokens, were formed with this suffix. This is

[88] The suffix *-ion* is present in this corpus under three different allomorphs, namely *-ion* (*opinion*), *-sion* (*vision, percussion*) and *-tion* (*computation, alteration*).

[89] The productivity of *-ion* is paralleled in other languages (Díaz, 2011, p. 131). According to Monge (1970, p. 969; 1977, p. 156) *-ion* was at the end of twentieth century the most productive suffix to form abstract action nominals in Spanish.

partly because the most repeated token (*motion*) is formed with this suffix. The motivation for this result is to be found extralinguistically: it is a result of the historical evolution of the discipline of astronomy. Immediately after the publication of *Principia Mathematica* (Newton 1687) in which the first two laws of motion were formulated, astronomers realized that the discipline was to change completely. In the two subsequent centuries –that is, the period covered in this corpus– scholars were busy in the application of Newton's theories and the calculation of the motions of the planets[90], hence the consistent repetition of the word throughout the sample texts in the corpus.

The next most productive suffix is *-ing*. This result is hardly surprising, as suffix *-ing* is the defining feature of one of the most studied typologies of nominalizations: gerundives[91] (Chomsky, 1970). Nominalizations ending in *-ing* sum up 543 tokens scattered in 93 different types. Figure 37 shows the chronological evolution of the two most productive suffixes in nominalization formation across the two centuries under study. There is an interesting tendency, as, whereas the number of nominalizations with *-ion* progressively augmented, nominalizations with *-ing* tended to decrease.

More specifically, the rate of variation of *-ion* nominalizations across the two centuries is 3.35% whereas that of *-ing* nominalizations is a negative 7.37%. This decrease in the rate of *-ing* nominalizations is particularly interesting because it contradicts the main tendency of slight, steady increase in the total number of nominalizations. In other ways, the evolution presented by *-ing* nominalizations is extremely irregular and it presents sharp variations: rates of variation are mainly over

[90] For more information, see chapter one.

[91] As explained in chapter two, *-ing* nominalizations have received a wide range of denominations by different, often generativist, scholars. Thus, Chomsky (1970) referred to them as *gerundives* and Quirk et al. (1985), as verbal nominalizations –as opposed to deverbal.

5% either positively or negatively between decades, whereas in *-ion* nominalizations the same rates of inter-decade variation hardly ever surpass 2%.

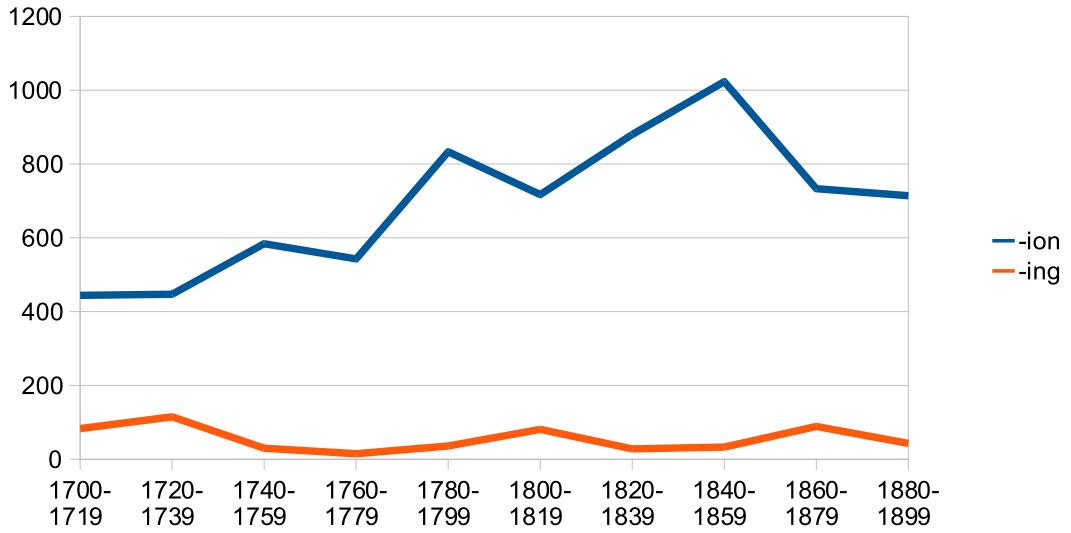


Figure 37: *Evolution in the use of -ion and -ing.*

One reason for this apparent unpredictability in *-ing* variation could be related to the decision of sampling only a subset of the total number of *-ing* nominalizations. After the disambiguation process, both occurrences of *-ing* and *-ion* forms have been excluded from analysis. However, given the fact that nominalizations are part of a continuum and that *-ing* nominalizations are closer to verbal realizations, many words ending in *-ing* have been left out from this study.

Concerning etymology, the origin of roots has been taken as another variable of study with the intention of analyzing whether the Romance or the Germanic element were productive sources. The results of this analysis can be seen in figure 38:

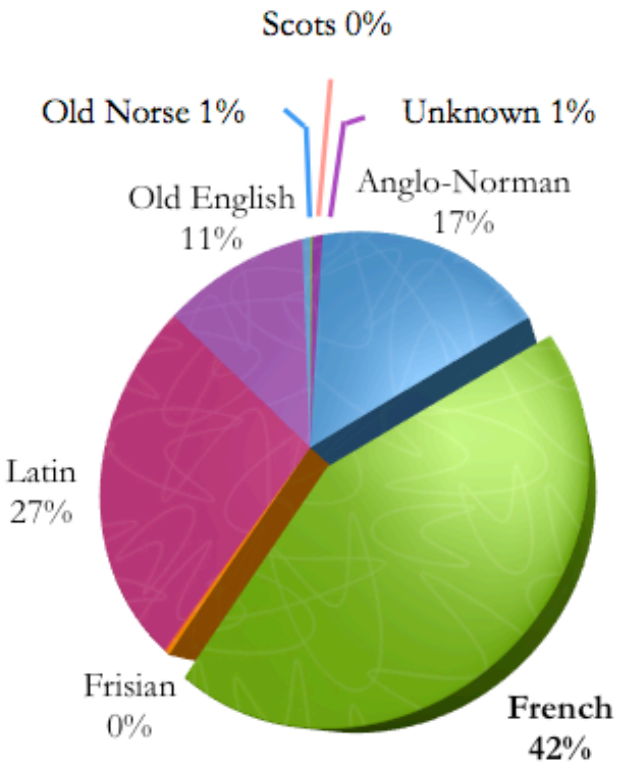


Figure 38: *Etymology of the root (types).*

The etymological distribution of nominalization roots clearly favors the Romance component of the language, with 563 Romance versus 79 Germanic types, that is a prevalence of 88% of Romance types. The difference is even bigger if we take in consideration the tokens because the percentage of Romance roots raises to 94% of the total. The composition of the Romance component in this subset feeds from different sources, namely Latin, French and Anglonorman[92]. French is the most productive source of nominalizations with almost half of the total, that is 3,938 tokens and it includes words such as *assistance, conjunction, apprehension* and *development.* Latin is the second most populous source of nominalizations. It includes 1,924 appearances and

[92] Establishing clear-cut distinctions between Anglo-Norman, French and Latin is highly debatable. However, the inclusion of these three sources as different languages has been made following the classification in the *OED*.

words such as *designating, assumption and assertion.* After Latin, Anglonorman, acknowledging 2,082 appearances is quite productive as well. Words falling in this group have a Latin/French origin and were introduced during the Norman conquest that took place in the eleventh century, which explains their assimilation to English standards. *Motion, sustentation, movement* and *revelation* are examples of words with an Anglonorman origin. To finish the group of Romance languages nurturing English nominalizations, the Italian component is timidly represented here with the word *management* being repeated three times in Olmsted's and Denison's texts.

Slightly less productive, the Germanic component is also present in the corpus. The core element of the language (OE) represents 11% of the total, that is 431 appearances and 72 different types, which is less than any of the major Romance groups but constitutes the 89% within the Germanic group. Nominalizations present in the corpus and evolving from OE may be *setting, beginning, reading* and *washing.* Old Norse is represented by five occurrences in three different types of words: *flattening*, *banging* and *happening.* Frisian is present in the corpus with only two types, *rising*, which appears forty-five times under three different spellings (*rising, rifing, and riseing*), and *ſouthing*[93], which appears only once. The last Germanic source for nominalizations is Scots with the sole appearance of *rigging*.

Interestingly enough, figure 39 bears a high similarity with the evolution of the most productive Romance and Germanic suffixes. Both Romance roots and suffixes tended to increase although the rates of change hardly ever surpass 2%, the biggest exception being the augmentation of 12.17% between the 1760s and the 1780s.

[93] The *OED* dates the first appearance of this nominalization in 1653 with the meaning of "the action of a celestial object in reaching the southern section of the meridian circle [...]. Also: apparent movement of a celestial object towards the south." The root of the word, *south*, was introduced before and has a Germanic origin.

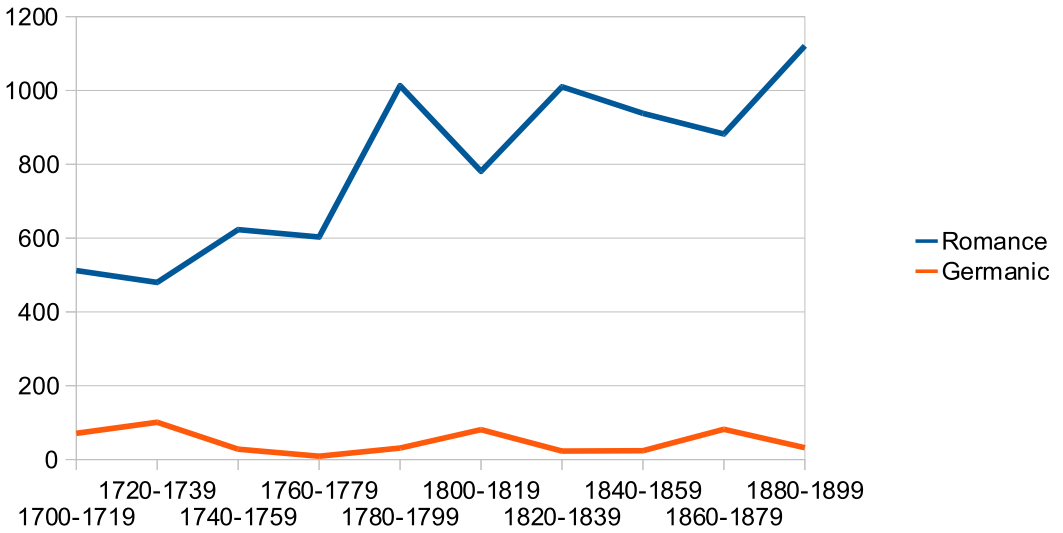


Figure 39: *Evolution in the use of nominalizations according to the origin of the root.*

As in the case of *-ing* suffixes, Germanic roots present high variability in their evolution and Romance are more stable. Even if their number is more reduced, the variation rates are usually over 10%, both positively and negatively. Most importantly, the tendency here is to diminish and the effect is especially evident if we pay a closer look at the first forty years of the eighteenth century and the two last decades of the nineteenth century.

Once the distribution and evolution of etymology of both suffixes and roots has been analyzed, the next step is to devote some attention to combinations of roots and suffixes from similar and different origins. As there are only two language families involved, there may be four resulting combinations, whose distribution is shown in figure 40:

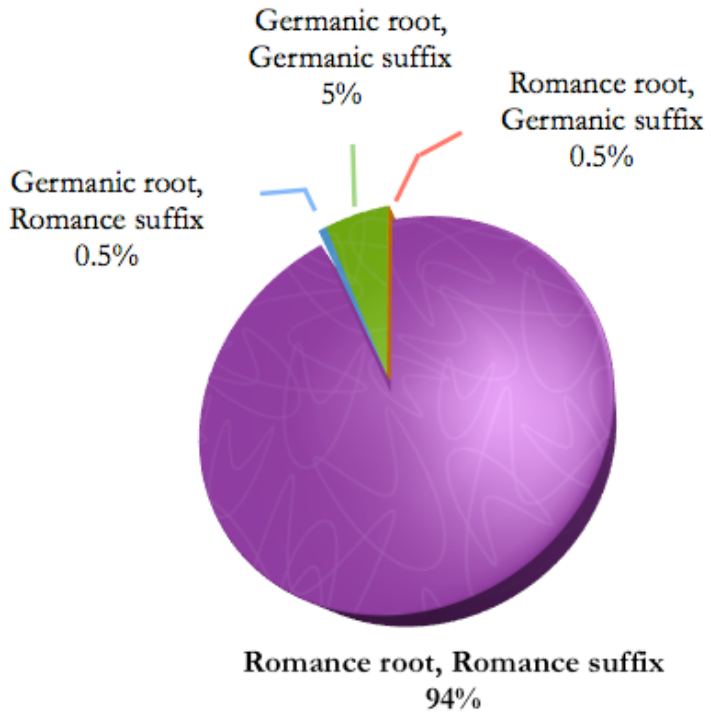


Figure 40: *Combination of roots and suffixes.*

The advantage of Romance words is outstanding. There are 7,894 words of this type, like *acceptance*, *correction*, *passage* or *arrival*, to name a few. The next most productive combination is that of purely Germanic words, with scarcely less than 500 tokens. This group includes words such as *feeling*, *beginning* or *meeting*. Blends of roots and suffixes with different origin seem to be less usual. Thus, the combination of Romance root and Germanic suffix presents only 63 tokens and has examples like *diminishing, establishing* and *wasting*. The opposite combination –Germanic root and Romance suffix– is extremely rare and there are only four examples of it in the corpus: *acknowledgement* (2 tokens), *amazement* and *wonderment*. The evolution of these four combinations is outlined in figure 41:

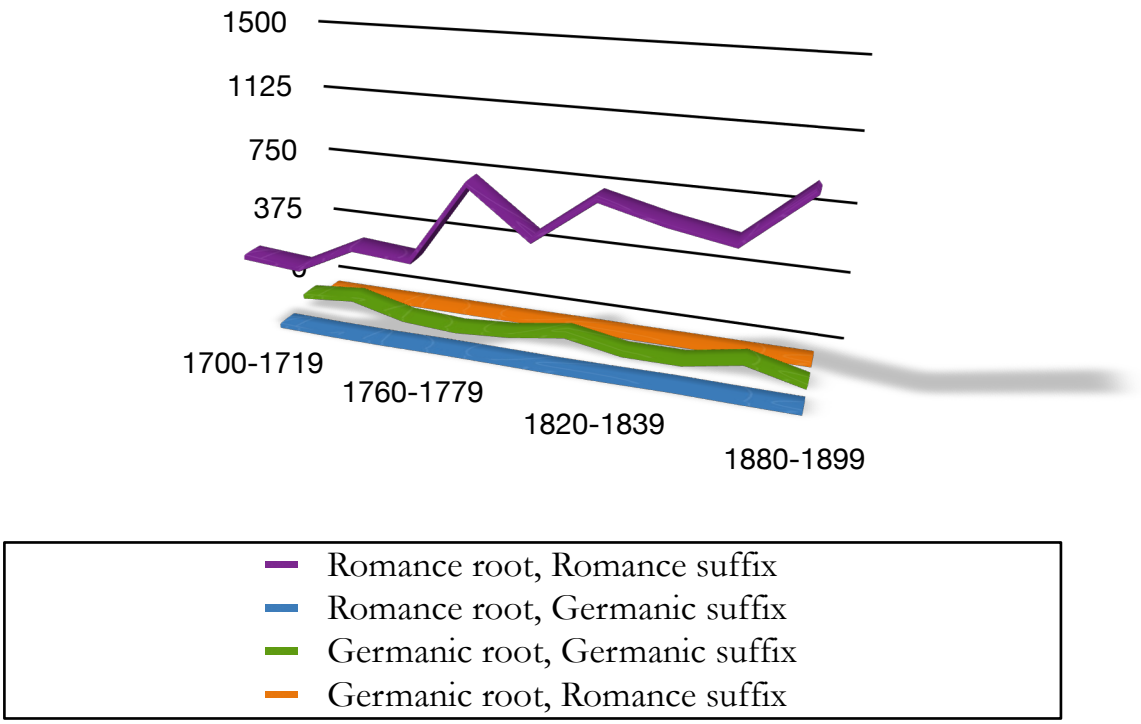


Figure 41: *Evolution of different root and suffix combinations.*

Complete Romance nominalizations (purple line in the graph) follow the general trend of nominalizations, that is a steady, progressive augmentation in frequency[94]. The mean frequency for this type of nominalizations in the eighteenth century is 640 and in the following century the frequency rises to 945, which can be considered a sound augmentation. Germanic nominalizations (marked with a green line) are the second most productive group but, in this case, their behavior is completely different, as there is a slight tendency to decrease. The mean frequencies for both centuries do not show major differences but it is true, however, that there is a peak in the first half of the eighteenth century of 175 nominalizations, unparalleled with the forty last years of the nineteenth century, which have less than half of those nominalizations. The remaining

[94] Actually, given the high number of nominalizations formed with Romance roots and suffixes in comparison with other combinations, it is very difficult to state that this type of nominalizations follows the general trend of nominalizations or, on the contrary, the general trend of nominalizations has been modeled upon this type of root and suffix combination.

hybrid groups present very similar trends, both showing consistent evolutions without consistent changes. Equally, hybrid groups are rare and they hardly ever surpass the barrier of five tokens per twenty years. Nominalizations with a Germanic root are a Romance suffix and even more scarce and it is even difficult to find any example of them in some time slots.

The next variable of study contrasts the origin of the root with the introduction of the nominalization in the language. This parameter of study is useful to assess whether nominalizations were mainly introduced in English as a result of the Scientific Revolution or, on the contrary, they were already common in the language and the development of the scientific register only meant an increase in their frequency. A reading of figure 42 supports the latter hypothesis.

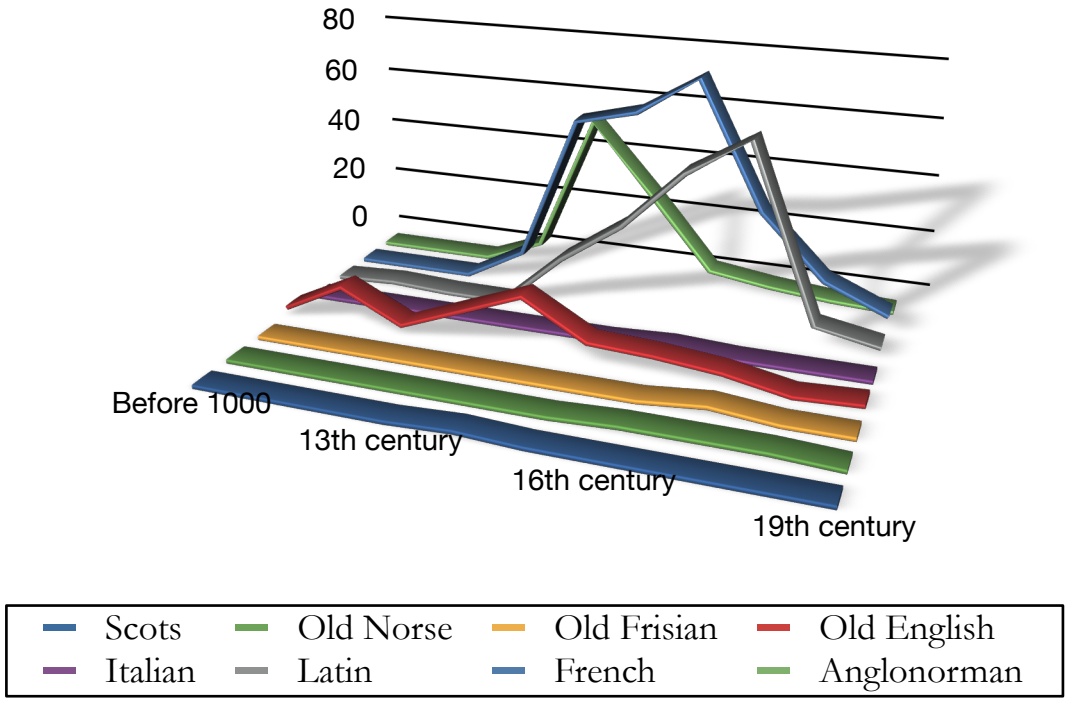


Figure 42: *Origin and year of introduction (types).*

According to this graph, nominalizations are indeed a clear model to exemplify the main movements of loanword acquisition in the history of English. The first peak in the figure indicates that Old English was the main source of nominalizations, or, in other words, that nominalizations were not being borrowed from other languages. In fact, until the thirteenth century, no other source of nominalizations is acknowledged, with the only exception of *procession*, coming from Latin.

After the Norman conquest, the following wave of loanwords came from Norman French. In the figure this is represented by the dramatic increase in the French, Latin and Anglonorman columns. This massive rise is not represented in the figure until the fourteenth century, which is related to the fact that nominalizations are indeed complex linguistic devices and they may express abstract ideas. This may definitely slow down their inclusion in the language, specially in cases in which, due to external reasons like a conquest, other more pragmatic, down-to-earth words may be prioritized when it comes to loanword acquisition[95]. The highest peak in the figure is marked by French nominalizations introduced in the sixteenth century. This century witnessed a massive input of loanwords into the language that were a result of several extralinguistic events, such as the introduction of the printing press[96], which increased literacy rates and called for the expansion of learned vocabulary, and the influence of the Renaissance[97], which fostered translations from classical texts and urged writers to refine and adopt new

[95] This is the case, for example, with words like *pork* (1215) and *duke* (1275? or earlier), both introduced into English as a result of the Norman conquest, which allude to common entities in the everyday life of the population of the time.

96 To be exact, the printing press was brought to England in 1476 by William Caxton, although it may be generally agreed that it was not until the sixteenth century that the effects of this historical event can be effectively seen in regular language.

[97] Here again, the slight inaccuracy in dates can be easily explained if we take into account that, although the Renaissance movement started in Italy in the late fourteenth century, it did not reach England until the end of the fifteenth century and the beginning of the sixteenth.

words so as to polish up English to set it at the level of classical-language standards. This last issue was also a blatant consequence of the Protestant reformation and the nationalist policies carried out during the Elizabethan reign, a period characterized by nationalist exaltation and pride in the English language. The amount of general borrowing during this period and the specific need for learned words characteristic of this century can account for the high figures of Latin and French nominalizations introduced in the sixteenth century that are reflected in figure 42.

Finally, the last peak in the figure, that is, nominalizations borrowed from Latin and introduced during the eighteenth century (marked with a grey line) is again a direct consequence of extralinguistic events. In this case, the Scientific Revolution was the main trigger for this result, as scientific texts were originally written in Latin and, while English was advancing positions in the writing of science, scholars, who were proficient Latin, incorporated new borrowings into English. This last part of the figure is, perhaps, the one that would differ the most from a figure measuring the origin of words and their introduction in English, as most of these Latin words introduced in the seventeenth century are closely related to new advances in astronomy, the scientific discipline under study. Of the 65 Latin types of nominalizations introduced in this century, a high percentage belong to the astronomy jargon and are accordingly marked in the *OED*. Some examples would include *circumgyration* (1603), *retrogression* (1604), *exposure* (1609), *gravitation* (1645), *scintillation* (1658), *deflection* (1665), and *collimation* (1687). An interesting feature that shows that English still retains its Germanic origin is the introduction of new combinations of specialized words with a Germanic root, such as *twirling* (1611), *banging* (1647) and *southing* (1653).

To sum up, it can be said that, even if their number and frequency of use rose in scientific register as a result of the Scientific Revolution, especially with words coming from Latin, the hypothesis that their use changed completely after the seventeenth century is inconclusive, as nominalizations were already a core element in the language. This Latin influence would ultimately result in a reinforcement of the Romance element, which is undoubtedly the predominant element contributing with both roots and suffixes to the formation of nominalizations. One plausible explanation is that this is a direct consequence of the effect of the Scientific Revolution. Apart from a revision of the methodology and the establishment of a new register, one of the main consequences of the Revolution was the abandonment of Latin as the language of science. Thus, in very little time, scholars started to write in English without major imposition. They were, however, also proficient in Latin. This may have influenced their linguistic choices and unconsciously they may have favored the Romance element which, simultaneously, resulted in the establishment of a trend.

4.1.2.2. The syntax of nominalizations

According to the traditional functionalist account of nominalizations (Banks, 2005a, 2005b; Guillén, 1998; Halliday, 1985, 2004; Ventola, 1996), which are mainly concerned with thematic structures and the advancement of discourse, nominalizations are most likely to be found in subject, subject complement or direct object positions. However, the analysis carried out for this study has showed, as can be seen in figure 43,

that nominalizations usually function as adjuncts or modifiers (55%). Of the 8,446 nominalizations studied, 4,675 were found functioning as modifiers, which represents more than 50%. It must be pointed out that this group includes not only nominalizations in PPs and NPs directly modifying the VP, but also those nominalizations modifying other nouns within an NP.

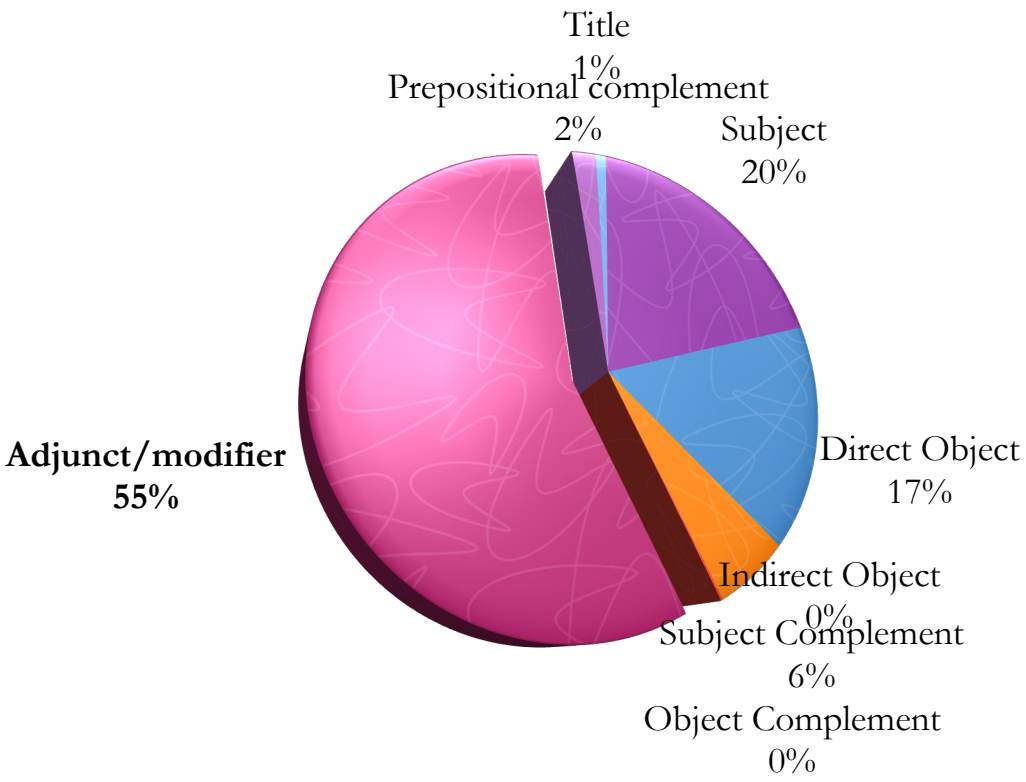


Figure 43: *Syntactic function of nominalization NPs.*

The explanation of the high frequency of nominalizations functioning as adjuncts can be found by having a closer look at deixis. Although some authors (Moreno

Cabrera, 1987, p. 65) restricted deictical functions to personal, spatial and temporal relations and word categories like pronouns, adverbs and conjunctions, other scholars (Graesser, McNamara & Louwerse, 2003) expanded the idea of deixis and deictical coherence to all types of relations in texts that may improve coherence. Even for them, the emphasis for deictical coherence usually falls on conjunctions and connectives, which are very common in expository texts because of their functionality when it comes to establishing relations between new ideas.

Cohesion, deictical coherence and nominalizations was a topic introduced by functionalists, which recalls on the definition provided by Halliday and Hassan (1976, p. 298), who understand cohesion as the type of unity "that exists between one part of the text and another." Concerning nominalizations, the functionalist tradition mainly focused on those cases in which verbs are turned into nominalizations fulfilling subject positions.

For them, nominalizations are a special type of lexical cohesion that repeat and summarize information, which facilitates the assimilation of the information. Even if this is true, the role of nominalizations functioning as adjunct has been usually neglected in spite of their high frequency. Nominalizations functioning as adjuncts are powerful deictical coherence devices. They establish temporal and spatial connections, draw connections between the processes in the text and serve as reference points in long, circonvoluted sentences.

In both (58) and (59) the underlying feature is closely related to deictical coherence of texts. In (58)

> (58) [Def]. XII. A Star riſes Heliacly, when after it has been in Conjunction with the Sun, it comes to be at ſuch a Diſtance from

> him, as to be ſeen in the Morning before **Sun-riſing** (Gordon,1726, p. 74; emphasis added).

the type of deixis is temporal, as the preposition *before* clearly indicates. In (59), however,

> (59) The point of the field then occupied by the intersection of the cross-threads is to be regarded as a fixed point of reference, and, as the telescope **revolves** from one position to another, the number of divisions of the limb which pass by this point will be the measure of the angular **motion** of the telescope (Chauvenet, 1871, p. 34; emphasis added).

the discourse deixis strategy is more elaborated because in some way, the nominalization is referring to a process of movement, which was introduced and explained before in the text in the verb *revolve*. In the following sentence, the verb is changed for the hyperonym *move*, turned into the nominalization *motion* and inserted in a context fulfilling a peripheral syntactic function in the postmodifying field of a subject complement NP. Apart from being economic and serving as a powerful cohesive device, nominalizations such as *motion* also convey some degree of discourse deixis which facilitates an economic information decoding process. In other cases, pronominal deixis works together with nominalizations functioning as adjuncts, as in (60),

> (60) I might content myself with barely referring **you** to that well-written article; but, as **you** may not have the work at hand, and would, moreover, probably not desire to read the whole article, I will abridge it for **your perusal**, interspersing some remarks of my own (Olmsted, 1841, p. 322; emphasis added).

where the relationship between the author and the intended reader is made not only through pronouns but it includes also the NP *your perusal*, which expands the referential cohesion to the process of *perusing* (*the book*). This has a direct impact on

coherence in discourse functioning in a particular communicative context. After all, this type of social deixis linking the author and the reader may be seen to add semantic continuity to text-type[98] and communicative context in which (60) is found.

After that of adjunct, nominalizations are likely to appear functioning as subjects (20%), which confirms the claims made by Halliday (1985, 2004) and other functionalists (Guillén, 1998; Banks 2005a, 2005b), who claim that nominalizations are key linguistic devices for information construction. According to these authors, nominalizations function as themes, which provide a backgrounding of the information presented previously in the text through verbal realizations and set the basis for the introduction of new information in the rheme. Following this reasoning, subjects are the logical function for nominalizations in scientific texts. The usual configuration of this type of nominalizations functioning as subjects can be found in (61)

> (61) The **conclusion** thence derived, that our globe weighs 4½ times as much as an equal bulk of water [...] was not very exact (Clerke, 1893, p. 321; emphasis added).

where the *conclusion* is introduced as subject of the sentence. It is completely reified to facilitate assimilation and presented as a mental process referring to a material process that is specified in a postmodifying field (*that our globe weighs 4½ times as much as an equal bulk of water*). In the second part of the sentence, the information expressed in the theme is negated (*was not very exact*). The nominalization in this sentence has served to summarize information and reify it as well as presenting ground for the advancement of discourse.

[98] Olmsted's letter tries to evoke direct speech and it is full of this type of cohesive elements. This intention to recreate direct speech is already made explicit in the title: *Letters on astronomy, addressed to a lady in which the elements of the science are familiarly explained in connexion with its literary history.*

The relatively high frequency of nominalizations functioning as subject complements (6%) is also a consequence of this particular way of constructing information in scientific texts. Halliday (1985, 2004) claimed that the trend for scientific register was that of combining two heavily modified NPs encoding a process linked by a semantically-emptied verb. There have been studies pointing out the progressive grammaticalization of verbs (Lehman, 2002). Apart from this new trend in language, copulative verbs have traditionally fulfilled this function, which explains the high frequency of nominalizations functioning as subject complements (6%). Generally speaking, the implications of nominalizations functioning as subject complements are similar to those functioning as subjects. In (62)

> (62) It is ſuppoſed by ſome Philoſophers, that a Comet may approach near enough to the Sun, not to be cold again in 50,000 Years. But this is too great a **Violation** of the Laws of Nature, for by its returning again, if it is but once in 500 Years, it would in a few Revolutions be vitrified, and at laſt diſſipated and loſt (Lacy, 1779, p. 22; emphasis added).

violation amalgamates all the traditional properties of thematic nominalizations. It encodes in one noun a process (*violating the Laws of Nature)*, which serves as a summary of what has been previously stated (that some philosophers have stated that a comet will pass very near the Sun). Lacy clearly disagreed with this and presented his judgment in the form of a reified nominalization, creating important ideological implications, as CDA scholars (Billig, 2008; van Dijk, 2008) have indicated. The effects of this nominalization are also achieved thanks to the demonstrative *this* functioning as the subject and linking the sentence where it appears with the preceding one. The role of

the copulative verb is equally important as it transfers the semantic force of the sentence to the nominalization in the subject complement position.

The grammaticalization of verbs in scientific register is not only restricted to copulative verbs and it applies also to other verbs, which take a nominalization as their direct object (17%). In these cases the semantics of 'process', typically associated to verbs becomes shared by the verb and its direct object nominalization in what could be considered a light-verb construction, as in (63):

> (63) During this last interval of seventy-six years, the science of mathematics [...] **has made prodigious advances**, more especially in its application to the laws of the celestial motions, as exemplified in the "Mechanique Celeste" of La Place (Olmsted, 1841, p. 334; emphasis added).

The transformation of the semantically full verb *to advance* in a collocation made up by a semantically-emptied verb that serves as a syntactic support to a process nominalization functioning as direct object does not respond to textual requirements and it is better associated with rhetorics and the creation of a codified scientific register. Light-verb constructions, are indeed a combination of a verb and a noun that form a single unit with no extra semantic value (Alonso Ramos, 2004, p. 17). The inclusion of a nominalization could be seen as a symbol that the author could achieve a high degree of abstraction and, additionally, that he knew the writing standards that belonged to the scientific discourse community. When choosing a longer expression writers neglected the principle of economy of language for the sake of style, and guild codifications. Concerning use, most of the nominalizations functioning as direct objects with a semantically weak verb found in the corpus are indeed collocations and, therefore, stylistic.

The function with the lowest frequency, titles (1%), is indeed very interesting because scholars have provided very little information about them. There are 63 nominalizations found in texts, which is a relatively high frequency if we take into account that titles are quite scarce in texts compared to other syntactic functions. It can be said that titles constitute the highest step in the reification process. Nominalizations indeed have some of the most important features titles need: they are concise and pack a great deal of information in just one word. At the same time they present a process in the same way a thing would be introduced, which facilitates the assimilation of the information by the reader. Finally, unlike verbs, which need a series of mandatory complements, all the pre- and postmodifiers of nominalizations are optional and may or may not be used by the author, as in (64),

> (64) **INTRODUCTION**. ASTRONOMY (astron, a star, and nomos, a law) treats of the Heavenly Bodies [...] (Steele, 1874, p. 13; emphasis added).

which is a fine example of complete modifier reduction. Cognitively, these nominalizations provide very effective guidelines to ensure a smooth information decoding by focusing all attention on the a highlighted process. Agents, circumstances and other participants in the process have been omitted to focus all attention on the process itself, which is presented as a reified entity. This factuality encompassed in titles has, in turn, very useful implications in the transmission and extension of knowledge. There does not seem to be a limit, however, to the type of verbal valencies included in the NP of the title. Hence in (65),

> (65) SECULAR **VARIATION** OF THE OBLIQUITY OF THE ECLIPTIC (Gummere, 1822, p. 235; emphasis added).

the postmodifying PP would be a direct object in case of verbal realization, whereas in (66)

> (66) CHAPTER XVII. **Universal Gravitation** and some of its effects (Gummere, 1822, p. 213; emphasis added).

the agent of *gravitation* appears as the adjective *universal* respecting the canonical SVO pattern. The conclusion after data analysis is that different syntactic functions have different functional implications in texts. The most surprising finding may be the high frequency of nominalizations in adjunct positions and their importance as discourse markers guiding the reader's decoding process and facilitating information assimilation.

4.1.3. Linguistic variables: the morphosyntax of nominalization NPs

This section is divided in two subsections. In section 4.1.3.1 the structure of the NP nominalization is analyzed. The emphasis here will lie on issues related to the description of the pre- and postmodifying fields in nominalization NPs. Special attention is paid to the role of possessives. It is important to describe all the elements in the nominalization to have an idea of how flexible nominalization NPs are in terms of constituents and how information related to the processes involved is codified in the phrase. At this level, it is also important to realize that in spite of all the verbal features nominalizations may retain at the semantic level, formally they are nouns and therefore can admit all types of nominal modification.

Section 4.1.3.2 deals with the functions fulfilled by nominalization modifiers regarding verbal realizations. In this section the focus lies in the functional description of elements appearing in the nominalization NP that encode elements that take place in the process encoded in the nominalization. As shown in previous sections, both in the pre- and post-modifying fields APs, possessive structures and PPs may hide information about the process. In case of a verbal realization, this information would be contained in obligatory verbal valencies or in optional adjuncts. However, in nominalizations all these elements become optional This definitely characterizes the information decoding processes implied in nominalizations as the focus lies on the process and on the additional information that has been included in the modifying fields.

4.1.3.1. The structure of nominalization NPs

Concerning the description of the modifying fields in NPs governed by nominalizations, it is logical to start with an analysis of the premodifiers. As figure 44 shows, there is a wide array of possibilities for premodifying elements.

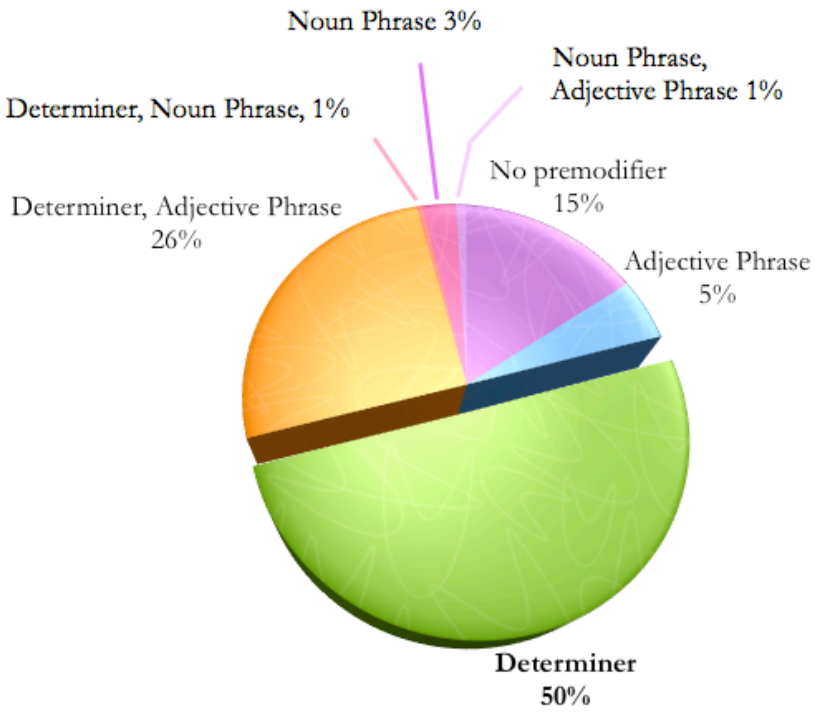


Figure 44: *Premodifiers to nominalizations*.

The most repeated premodifiers are determiners, which represent more than 75% of the total if we also take into account those examples in which the determiner is combined with an AP or an NP, as in (67) and (68). In (67)

> (67) **The greateſt elongation** of an inferior planet is when a line TE, drawn from the earth at T, through the planet at e, is a tangent to the orbit of the planet (Adams, 1777, p. 8; emphasis added).

the adjective *greatest* modifies the reified nominalization *elongation*. In other words, *elongation* here is presented as an entity and then a superlative is attributed to it. In (68), on the other hand,

> (68) With like disregard of the effect due to distance, the canals to the east of the Ganges showed better at **the November presentation** (Lowell, 1895, p. 113; emphasis added).

the noun premodifying the nominalization *presentation* is actually introducing temporal information about the process. Premodification through NP (3%), however, is almost restricted to Saxon genitive constructions, which normally indicate the agent of the process, as in (69)

> (69) [...] this is **Hartſdorf's innovation**, and it is very pardonable in reſpect to Schiller's [...] (Hill, 1754, p. 3; emphasis added).

although there are some cases in which the agent is introduced as an NP without a trace of possession, as in (70)

> (70) But, more than all this, a very considerable part of the 1 foot presently being removed off the country in 6,000 years consists of the loose materials belonging to the glacial epoch, such as sands, gravels, and boulder clay, which are being swept off the surface by rain and **river action** (Croll, 1889, p. 47; emphasis added).

Regarding possible functions in relation to the premodifying elements and the process expressed in the nominalization, whereas NPs and Saxon genitives tend to indicate an agent, as in (69) and (70), the relation is more flexible in the case of adjectives: we can find adjectives that would function as either subjects or adjuncts in case of verbal realizations. In (30),

> (30) For mankind muſt have made conſiderable advances in aſtronomical learning, before they could ſo far diſengage themſelves from the prejudices of ſenſe and **popular** opinion, as to believe in a doctrine ſo ſublime, and remote from **vulgar**

> **apprehenſion**, as that which the moderns have now firmly eſtabliſhed (Bonnycastle, 1786, p. 55; emphasis added).

the adjective *vulgar* –together with the adjective *popular* appearing in the previous line– refer to people and *vulgar apprehension* could be rewritten as *what people apprehend*. In (71), on the other hand,

> (71) The inequality called the Acceleration of the Moon by which her velocity appears subject to continual increase, and her period to **continual diminution**, has been found by Laplace to be a Secular equation, [...] (Gummere, 1822:, p. 227; emphasis added).

the adjective *continual* conveys the idea of a circumstance about the process expressed in the nominalization (the period of the Moon diminishes continually).

Lack of premodifier (15%) is also quite productive and it seems that it is not related to the position and/or function of the nominalization in the sentence. It could be expected that nominalizations within less wordy phrases should be found fulfilling peripheral functions –such as adjuncts or modifiers within other sentences–, where they would have a referential role. However, it is easy to find instances of non-premodified nominalizations functioning as modifiers, as in (72)

> (72) This is alſo the caſe in what they call conſequent angles, and the term equal is uſed, by way of **diſtinction**, between the one and the other kind; though ſome call both conſequent (Hill, 1754, p. 15; emphasis added).

In (72) the nominalization is fully reified and presented as a circumstance in another process. Lack of premodification is also widely found in light-verb constructions featuring a nominalization as the direct object, as in (73)

(73) By making **Tryal** it will appear, that the Square of Saturn's Periodical Time, is to the Square of Jupiter's, as the Cube of Saturn's Diſtance from the Sun, is to the Cube of Jupiter's Diſtance (Gordon, 1726, p. 112; emphasis added).

Regarding the existence of determiners premodifying the nominalization almost all the determiners listed by Quirk et al. (1985) can co-occur with a nominalization, as can seen in figure 45. This indicates that nominalizations are formally assimilated as nouns even if they may maintain some verbal semantic features.

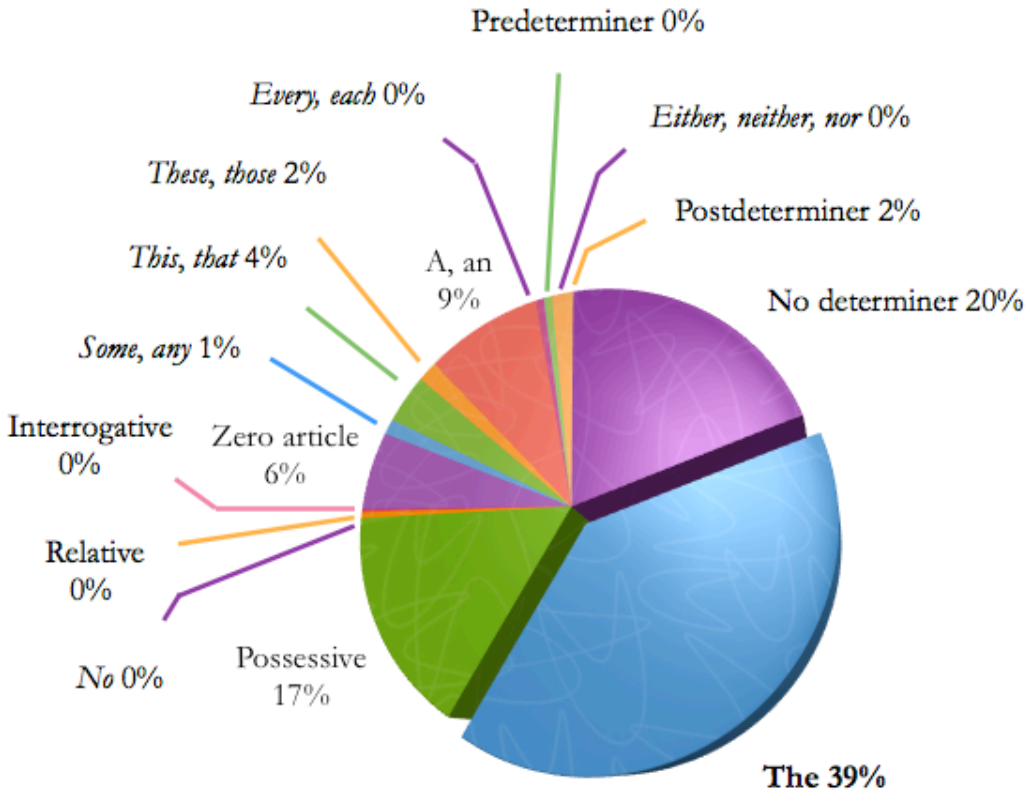


Figure 45: *Determiners in nominalization NPs*.

As might be expected, the article *the* is the most repeated determiner (39%) and *a/ an* (9%) and the zero article (6%) are reasonably common as well. The high frequency of possessive determiners may be explained by their referential quality which, in most

cases, links the process expressed in the nominalization with an agent or a result that appears in previous lines. The flexibility for text-construction they allow when combined with nominalizations renders them worthy of a deep analysis. The only determiners that have not been found in combination with a nominalization are those ending in *-ever* and *enough* but this fact may be related to their low frequency of use and the size of the corpus rather than other co-occurrence restrictions. The rest of the determiners described by Quirk et al. (1985) can all be combined with nominalizations. In (74)

> (74) Nor was it certain that there was not [...] one or more planets beyond Saturn, **whose attractions** might likewise influence the motions of the comet (Olmsted, 1841, p. 331; emphasis added).

attractions is premodified by an interrogative determiner that links the clause to the previous sentence and provides ground to place the nominalization in the subject position, that is, the point of departure for thematic structure in the relative clause. In (75)

> (75) In cristalography, we know not **what modifications** may take place, when the facts of dimorphism and plesiomorphism shall have been collected and classified, and the laws they follow discovered (Garland, 1838, p. 123; emphasis added).

the nominalization is accompanied by an interrogative. This is a quite uncommon construction, and so is premodification by *neither/nor,* as in (76):

> (76) Since, as they said, the circle is perfect, is the most beautiful figure in nature, has **neither beginning nor ending**, therefore it is the only form worthy of God,[...] (Steele, 1874, p. 25; emphasis added).

Indeed, there are only four occurrences of this determiner. Concerning predeterminers the only subtype not appearing in this corpus is that of fractions, whereas the most common postdeterminers are numerals, as shown in (77):

> (77) [...] but, as this mark will generally be found between **two divisions**, some additional means are required for measuring the fraction of a division (Chauvenet, 1871, p. 30; emphasis added).

Judging by the results obtained after corpus exploitation, it seems sound to state that there are no restrictions concerning what determiners can be found in a nominalization NP. In fact, the combinations of nominalizations and the different types of determiners seem to adjust to the general frequencies of use in English. This supports the idea that nominalizations display all nominal features to the full extent.

Possessives are the second most recurrent determiner premodifying nominalizations (17%). They are indeed a great source of information of pragmatic and functional features that can help draw divisions among the different types of nominalizations. In this study not only possessive pronouns but also other possessive constructions, namely Saxon genitive and *of*-constructions are included following the presumption that in most cases these possessive constructions would mirror a verbal realization, as in (78):

> (78) CHAPTER [I]. [...] —THE SUN'S **MOTION** AMONG THE STARS (Mitchel, 1860, p. 1; emphasis added).

The mirroring of a verbal structure in (78) is clear, as judged by the position of elements simulating a SVO pattern. A similar effect is achieved in (79)

(79) The path deſcribed by a planet in **its motion** round the ſun is called its orbit (Adam, 1777, p. 4; emphasis added).

where the possessive *its* serves as a cohesive device that links the nominalization with its immediately preceding agent (*a planet).* The distribution possessive use can be seen in figure 46.

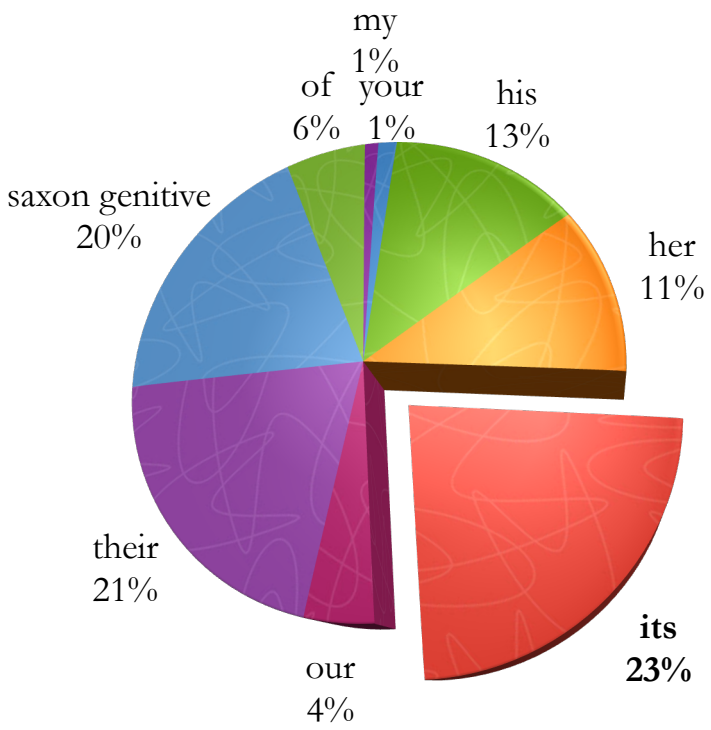


Figure 46: *Possessive constructions in nominalization NPs.*

Third person pronouns are the most frequent type of possessives in this study (68%). One of their main features is that, unlike other possessives, their referent is present in the text, many times serving as a substitute and adding lexical cohesion to the text. This is especially evident in those cases where the referential NP is the subject of a verb turned into a nominalization as in these cases the combination of a third person

possessive determiner and a nominalization is in some way a reduced clause mirroring the structure of a subject and its verb. In example (80)

> (80) **All theſe vaſt orbs** were imagined to move round the earth once in twenty-four hours and alſo in certain ſtated or periodical times agreeably to **their** annual changes and **appearances** (Bonnycatle, 1786, p. 57; emphasis added).

all these vast orbs appears as the subject at the beginning of the sentence. Then, it is turned into the possessive *their* and combined with the nominalizations *changes* and *appearances*. The clause *the vast orbs appear and change annually* has thus been converted into an NP to make it fit into the paragraph.

The differences of use between masculine and feminine possessive determiners can be explained if we take into account determiners may refer to either animate or inanimate entities. In the case of scientific astronomy texts from this period, the possible referents for the masculine possessive may be either animate (43%) or inanimate/ celestial objects (57%). In the case of animate referents, the biggest group is made of other scientists, represented by 36% of the total of masculine third-person possessives. A wide array of both celebrated and less known astronomers, such as Ptolemy, Copernicus, Brahe, Galileo, Kepler, Newton, Halley and Schiaparelli, are present as referents, as in (81)

> (81) Though **Copernicus** thus simplified so greatly the Ptolemaic theory, he yet found that the idea of circular orbits for the planets would not explain all the phenomena; **he** therefore still retained the "cycles and epicycles" that Alfonso had so heartily condemned. For forty years **this illustrious astronomer** carried on **his observations** in the upper part of a humble, dilapidated farm-house, through the roof of which he had an unobstructed view of the sky (Steele, 1874, p. 24; emphasis added).

where the reference to the well-known astronomer Copernicus is continuous not only by the possessive premodifying the nominalization (*his observations*) but also by lexical (*this illustrious astronomer*) and pronominal (*he*) means. There is also some degree of corpus intertextuality, as two referents are other *CETA* writers[99], as in (82),

> (82) This opinion has been supported and argued with a considerable degree of plausibility, by **[Dr]. Brewster** — grounding **his arguments**, very ingeniously, upon the phenomena we have just noticed as being peculiar to them (Phillips, 1817, p. 69; emphasis added).

which reminds us that authors are not idle entities and that scientific communities play an important role in science formulation and transmission[100]. Nevertheless the range of animate referents is not limited to others scientists. In (83)

> (83) But the earth meanwhile turns round at the rate of 15° per hour; and since **the observer** is unconscious of **his** own **motion** of rotation, it results that the plane of vibration of the pendulum appears to revolve at the same rate in the opposite direction (Loomis, 1868, p. 32; emphasis added).

the referent, and agent, of *his motion* is *the observer* himself. With a less eye-catching 7% of the total, other animate referents for third-person masculine possessives include

[99] Margaret Bryan (1797, p. 115) alluded to James Ferguson (1710-1776) and William Phillips (1817, p. 69) did so with Sir David Brewster (1781-1868).

[100] Needless to say, reference in third-person singular possessive pronouns modifying nominalizations formed via suffixation is only the tip of the iceberg concerning how guild codification has affected scientific English. However, the fact that even in such a specific domain, two examples can be found is indeed an indicator of the importance of this matter.

God[101], the *King of Britain* and *man* (82), as a rhetorical device[102]. Among the most repeated masculine referents for third-person possessives is *the Sun*[103] with 25% of the total, as in (85).

> (85) [...] and as **the Sun** in **his** diurnal **Motion** always moves parallel to the Equinoctial he muſt be longer above the Horizon than below (Charlton, 1735, p. 44; emphasis added).

Other planets and satellites are also present here, either with the general *any planet* or with their proper name[104], as in (86):

> (86) **Jupiter**, the largeſt of all the planets, is ſtill higher in the ſyſtem than Mars. [...] **His** diameter is computed to be 89,000 miles, and by a prodigiouſly rapid motion upon **his** axis, **he** performs **his diurnal rotation** in nine hours and fifty-six minutes (Bonnycastle, 1786, p. 37; emphasis added).

where *Jupiter* is the agent of the nominalization *motion* that is repeated in multiple forms (*he, his*) across the two sentences. Regarding the types of processes involved in the nominalizations premodified by masculine possessives, more than 75% of the total are material processes, especially, but not restricted to, those cases in which the possessive is inanimate and can function as the *actor* of the *action* or *event* like (83),

---

[101] The study of the heavens was for many centuries linked to religion and theology.

[102] The texts including *man* as a referent are textbooks, and the intention of the authors in this linguistic choice was undoubtedly, that of making the text easier to the reader by adopting a more comprehensible, closer to the point of view of the reader.

[103] *The Sun* can also be premodified by the possessive *its*, as in (84)

> (84) The Other on the South ſide of the Equinoctial is called the Tropick of Capricorn in which Point **the Sun** hath **its** greateſt Southern **Declination** making our ſhorteſt Day and longeſt Night [...] (Morden, 1702, p. 18; emphasis added).

[104] In this specific case, only the planets Jupiter and Mercury are explicitely named.

(85) and (86). It is also possible to find mental and verbal processes, as in (81) and (82) but these are restricted only to those cases in which, according to Halliday (1985) and Ravelli (1988) the possessive is an animate, human-like entity.

In the case of the feminine possessive socio-historical evidence related to the lack of female astronomers explains why all feminine singular possessives in this corpus refer to inanimate entities. Among these, the most frequent is *the moon* with almost 75% of the total, as in (87)

> (87) Let us abſtract at preſent from the inclination of **the Moon**'s orbit and ſuppoſe that **ſhe** performs **her** menſtrual **revolution** about the Earth in the Ecliptic and likewiſe that **her motion** in it is equable (Costard, 1767, p. 282; emphasis added).

where, as what happened in (81) and (86), there are multiple rewritings of the referent *motion*, not only as the possessive *her* but also as the personal pronoun *she*. Other feminine celestial objects are *Juno*, *Venus* and *the Earth*, as can be seen in (88).

> (88) [...] ſo that the points enfma, whilſt **the earth** performs **its revolution**, may run through the orbit of the planet (Adams, 1777, p. 10; emphasis added).

As a curiosity, there are two examples of nominalizations premodified by inanimate entities personified under feminine gender. These are *Athens* and *Nature.* Concerning types of processes involved in the nominalizations, all instances include, as might be expected, a material process because of the lack of female astronomers.

On the other hand, the high frequency of the determiner *it* is explained by the fact that astronomy is mainly concerned with inanimate celestial objects, like planets and stars. From this fact it is deduced that all processes involved in the nominalizations are,

as in the case of feminine possessives, material processes. Another distinctive feature of nominalizations premodified by *its* is that there is a tendency that the possessive would be the object –and not the subject– in case of a verbal realization. In (29)

> (29) While the general aspect of the planet reminded him of that of **Mars**. [...] but the difficulties in the way of **its observation** are enormously enhanced by **its** constant close **attendance** on the sun (Clerke, 1893, p. 304; emphasis added).

there is an example of the possessive *its* functioning both as Object (*its observation/we observe Mars*) and subject (*its attendance/Mars attends on the Sun*).

The possessive *their* also has a very high frequency rate. The most common syntactic relationship between the possessive and the noun is that of agent of the process. In (89),

> (89) **The planets** are diſtinguiſhed from the fixed ſtars, by **their motion**, and the ſteadineſs of their light (Adams, 1777, p. 3; emphasis added).

instead of choosing a verbal realization that would rephrase the whole sentence, the possessive indicates the syntactic relationship with the nominalization and, at the same time, refers directly to *the planets*. Regarding types of processes involved, there is a slight tendency to find more material processes, although there are also frequent examples of mental and verbal processes referring to human or human-like agents.

First person possessive determiners prove specially useful for rhetorical concerns. Whereas the first person singular determiner *my* serves to highlight the individuality of the writer, as in (90)

> (90) [...] a state of things which continued up to the close of **my observations** toward the end of November (Lowell, 1895, p. 115; emphasis added).

the plural form *our* can be classified as a type of guild codification as, in most cases, it refers to the whole scientific community as in (91)

> (91) [...] as in revolving in very long ellipses, they are sometimes too remote for **our inspection**, their greatest distance from the Sun being far beyond the orbit of the Georgium Sidus, as these bodies are not much larger than our Moon (Bryan, 1797, p. 94; emphasis added).

although it may also refer to whole population of the earth as in (92)

> (92) The obvious conclusion from this is, that [...] they are all performed nearly in the plane of the ecliptic, –that plane, namely, in which **our** own **motion** about the sun is performed (Herschel, 1833, p. 234; emphasis added).

or it can even be a way of referring to the author himself as in (93):

> (93) For theſe Reaſons, it is neceſſary, in the Beginning of this **our Aſtronomical Undertaking**, to ſpeak of our Earth itſelf, [...] (Whiston, 1715, p. 1; emphasis added).

Whereas third person possessives are related to textual features and lexical cohesion and first person possessives depend on stylistic concerns, the use of second person possessives is dictated by register and text-type. In some way, the use of *your* is also a stylistic choice but it is only found in learner's texts, namely dialogues, textbooks and lectures. Harris' *Astronomical dialogues between a gentleman and a lady* (1719) is the text with the highest frequency of second person possessives. In (94)

> (94) **YOUR Objection** is juſt, ſaid I, Madam if you conſider the thing after the Sun was actually Riſen and juſt before his Setting (Harris, 1719, p. 28; emphasis added).

the use of *your* is merely a rhetorical device and it refers to an imaginary lady the author created as part of his narrative.

In the case of Bryan's lecture the use of *your* is determined by the pedagogical aim of the work and it refers to the intended reader, who is supposed to be a learner. In (95)

> (95) Having premised so much I shall content myself with bringing **you** acquainted with the nature of those things mathematicians apply to the sciences they cultivate in order to ascertain the real sizes distances &c. of the heavenly bodies without perplexing **you** with theories too profound for **your contemplation** (Bryan, 1797, p. 104; emphasis added).

it can be easily noticed that Bryan addressed directly to the reader by using first and second person pronouns, in an attempt to imitate the direct speech that was used in the lecture that served as a basis for this publication.

Saxon genitive constructions are also common (20%). In most cases, the syntactic relation between the genitive NP and the nominalization would be that of a subject and its verb, as in (96).

> (96) This, however, is not all. **Schröter's description of Linné**, as seen by him November 5, 1788, tallies quite closely with modern observation (Clerke, 1893, p. 328; emphasis added).

Clearly, this structure emulates the relationship established by third-person possessives and nominalizations. In the case of Saxon genitive constructions, however, the need for specification of the possessor makes necessary the expansion of the premodifying possessive field. There are as well some instances like (97),

> (97) If we abandon the earth as a center of the planetary motions, it cannot admit of a **moment's hesitation** where we should place that center with the greatest probability of truth (Herschel, 1833, p. 246; emphasis added).

where there is no trace of *subject/verb* relationship between the genitive NP and the nominalization. In (97), *moment* comes to express some type of temporal meaning to the process expressed in *hesitation* in a sentence that could be rewritten as *we cannot hesitate for a moment*.

The timid 6% of nominalizations with a postmodifying possessive *of* construction was highly expectable. The expression of possession with this type of construction is more difficult to find combined with nominalizations and in most cases it is difficult to find traces of syntactic relationships between the elements. In (98)

> (98) PHYSICAL **CONSTITUTION OF THE SUN** (Steele, 1874, p. 61; emphasis added).

there are serious doubts whether the relationship between *constitution* and its postmodifier could be that of an agent and a process, that of a process and an object or there is no syntactic relationship involved at all. It might be consequently said that possessive *of constructions* are closer to term nominalizations, that is, fully reified nominalizations. In this type of possessive structure, nominalizations have virtually lost

all the verbal and semantic properties they might have retained and are closer to nouns at all levels. In (99),

> (99) [...] and that there is nothing elſe wanting to the eſtabliſhing that Motion, and unto the thorough **Conviction of the moſt obſtinate Adverſaries**, but that a Parallax of theſe Stars might be perceiv'd according to the diverſe Poſition of the Earth in its Annual Orbit: [...] (Whiston, 1715, p. 29; emphasis added).

on the other hand, there is still an underlying syntactic relationship between *conviction* and the agent in the process, *the most obstinate adversaries*.

Despite their frequency is not extremely high, data analysis has showed that the role of possessives can be very complex. It is possible to obtain a great degree of information about rhetorics, sociolinguistics and text-type by studying the relationship between nominalizations and the possessive structures modifying them. The most important point, however, may be to realize how possessives and nominalizations may work together to create a web of textual coherence, not only inside the text but also in relationship with the audience. In some way, possessives and nominalizations share the property of maximizing their functionality in organizing the information in the text while remaining extremely economic.

The existence of postmodifiers is more reduced than that of premodifiers. The rate of postmodifier presence is 50-50. In figure 47 the different types of postmodifiers appearing next to nominalizations can be easily seen. As in the case of premodification, there seems to be no restrictions at the level of what constituents can postmodify nominalizations. Thus apart from PPs, nominalizations can be modified by either NPs, relative clauses or non-finite clauses.

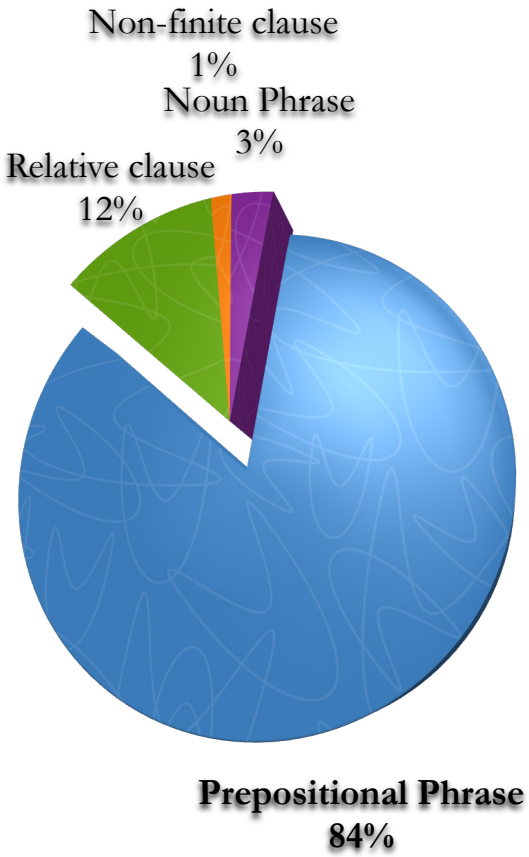


Figure 47: *Postmodifiers in nominalization NPs*.

Clausal postmodifiers represent only 13%, maybe because of the grammatical complexity they entail. They include relative clauses, nominal *that*-clauses and *to*-infinitive clauses, that is, all the possible options for nominal postmodification, which reinforces again that nominalizations share all the properties of common nouns at a structural level.

Relative clauses are the most common type of clausal postmodification. With almost 500 occurrences, they represent 12% of the total. Both defining and non-defining relative clauses are widely used. The number of defining relative clauses, like (100)

> (100) To account for these phenomena, the **explanation that at once suggests itself** is, that a direct transference of water takes place over the face of the planet, and that the canals are so many waterways (Lowell, 1895, p. 115; emphasis added).

is slightly smaller and in most cases they are reduced relative clauses as in (101)

> (101) The **conclusion drawn from a full examination of the subject** is, that the duration of the earth's rotation may be regarded as perfectly unchangeable (Gummere, 1822, p. 236; emphasis added).

following the author's wish to cut down on excessive wording. In (102),

> (102) It is as easy for Him to establish circulation and nutrition in material structures, as cohesion and crystallization, which we must suppose the planetary masses to possess; or **attraction** and inertia, **which we know them to possess** (Whewell, 1858, p. 57; emphasis added).

on the other hand, is a canonical example of non-defining relative clause, where the information provided in the subordinate clause does not restrict the set of elements present in the head. The less populated subgroup within clausal postmodifiers is that of *to*-infinitive clauses, which can be seen in (103).

> (103) In general they conform to their representations, and **failure to do so** is explicable not only by errors of observation, but by certain other facts (Lowell, 1895, p. 114; emphasis added).

Concerning nominal *that*-clauses, it was expected to find a good number of them because, as scholars (Burton-Roberts 1998, p. 204; Downing, 1992, p. 463) repeatedly

remark this type of modifying clause emulates the relationship between the verb and its complements within the VP and tends to occur in deverbal NPs. They may prove specially useful in verbal processes where direct speech is to be reported, as in (104):

> (104) Another answer is found in the **suggestion**, **that the crust of the earth beneath the deepest part of the ocean may be equally thick as in other parts, but more depressed or indented** (Bradford, 1845, p. 92; emphasis added).

In this particular example it can be seen how nominalizations can also have the same effect that passives in the sense that valency reduction deviates the attention from the agent to the object itself. However pertinent that-clauses may be in relation with nominalizations, only 57 examples of were found.

Concerning non-clausal postmodifiers, nominalizations postmodified by an NP are quite scarce and, according to the results, quite specific of a particular type of nominalization. Indeed NP postmodification may be one of the defining features of term nominalizations. The idea of whether the head of the NP is the nominalization or the subsequent nominal group can also be highly debatable. It is true that in these nominal compounds the semantic force lies on the nominalization but morphologically it would be also possible to think that the second noun is the phrase head. However, since this study is concerned with nominalizations, the analysis that was preferred to elicit simplicity is that of considering postmodifying NPs as appositive elements fulfilling the function of *naming*, according to the definition provided by Downing (1992, p. 463), who defines a nominal qualifier as a "unit that has the same reference." In (105),

> (105) When the Moon is in her greateſt north declination at M, the higheſt elevation G under her, is on the Tropic of Cancer and the

> oppoſite **elevation F** under her, is on the Tropic of Capricorn and theſe two elevations deſcribe the Tropics by the Earth's diurnal rotation (Ferguson, 1753, p. 153; emphasis added).

the group *elevation F* represents a typical case of nominal postmodification. In most examples found in the corpus, the postmodifying noun is usually a letter or a combination of letters, which reinforces the idea that these nominal groups are indeed modifying the head.

Of those instances of postmodified nominalizations, 3,525 occur with a PP. This high frequency prompted the need for a closer examination of their function. A small percentage of cases (7%) showed no connection with the expression of a process and are more prone to appear with those nominalizations that have acquired more nominal features, as in (106).

> (106) [...] nay, ſome to have their **Revolutions of Appearing and Diſappearing in exact ſpaces of time**; [...] (Morden, 1702, p. 38; emphasis added).

In other cases, such as (107)

> (107) PHYSICAL **CONSTITUTION OF THE SUN**. —Of the **constitution of the sun**, and consequent cause of the solar spots, very little is definitely known (Steele, 1874, p. 61; emphasis added).

the postmodifiying PP can be a possessive construction.

However, apart from this marginal category, most postmodifying PPs indicate in one way or another the elements that take place in the process expressed in the

nominalization. The functions of these elements within the nominalization NP are the subject of study of next section.

4.1.3.2. The functions in nominalization NPs

In this section the elements appearing in the NP that encode elements about process encoded in the nominalization are analyzed. Although nominalizations are not understood here as verbal transformations, in this section they are many times compared with them. The aim of this section is to analyze in what ways nominalizations have more flexible ways of organizing information related to processes to maximize text coherence, cohesion, facilitate the advancement of discourse and ultimately help guide the information decoding process. To this end, figure 48 shows the inclusion of transferred verbal valencies in NPs governed by nominalizations, both in the pre- and in the post-modifying fields.

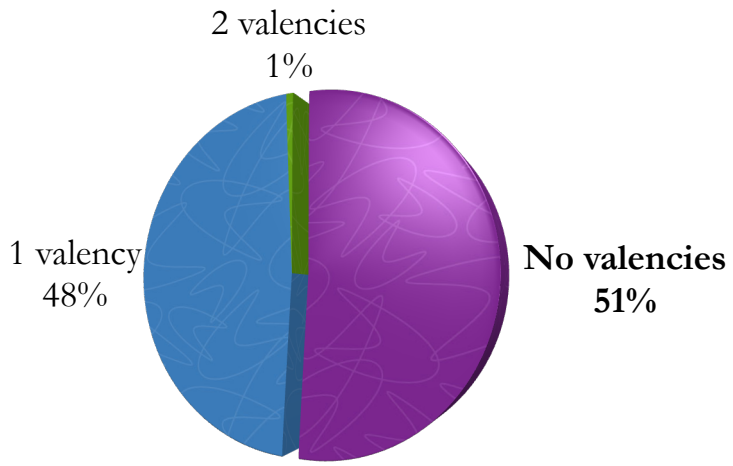


Figure 48: *Verbal valencies transferred into nominalization NPs.*

In figure 48 it can be seen that most nominalizations appear without any verbal valency. There may be other words accompanying the nominalization but the function of these words may have no connection to verbal patterns whatsoever. In (108),

> (108) The following **quotations** furniſh us with a remarkable inſtance of this, and will ſerve to give us a view of the ideas, which he came to entertain upon theſe ſubjects (Wilson, 1773, p. 4; emphasis added).

both the determiner and the adjective accompanying the nominalization cannot be re-accommodated in a verbal reading and consequently they are not considered verbal valencies. The distinction between verbal adjuncts and nominal modifiers is, however, a difficult task. The fuzzy interconnections between morphology, syntax and semantics are clearly exposed in (109). In this example the premodifying AP could be possibly rewritten as a verbal modifier *X determined the mass of the earth accurately.* It is also possible that *determination* once nominalized and reified acquired all noun features and consequently could be given the quality of *accurate, more accurate than* or *the most accurate (of all determinations of the mass of the earth that have been made by scientists).* The aim of this debate is not to reach a solid conclusion about the origin of *determination* and all its modifiers, but rather to show that meaning construction is extremely subtle and on occassions no clear dividing lines among semantics, morphology and syntax can be drawn.

> (109) If the mass of the earth be denoted by 1, the mass of the moon, according to **the most accurate determination**, is 1/68.50 (Gummere, 1822, p. 217; emphasis added).

One-valency transference (48%) is almost as frequent as non-valency inclusion (51%). Falling into this group, it is possible to find both pre- and post-modifying subjects as in (110) and (111)

> (110) The direction of the meridian may be secured at every instant by observations, and although local difficulties may oblige us to deviate in **our measurement** from this exact direction, [...] (Bradford, 1845, p. 90; emphasis added).

> (111) The fixed ſtars are diſtinguiſhed from the planets by being more bright and luminous, and by continually exhibiting that appearance which we call the **ſcintillation, or twinkling of the ſtars** (Bonnycastle, 1786, p. 44; emphasis added).

and also direct objects, as in (112)

> (112) CHAPTER IV. FURTHER **STATEMENT OF THE DIFFICULTY** (Whewell, 1858, p. 48; emphasis added).

Despite two-valency transformations are substantially less frequent (1%), they are also worthy of mention. Their scarcity is due to the fact that not all types of nominalizations can admit long NPs because of the textual functions they have acquired. This is the case of term nominalizations, where deictical functions directly compress the size of the phrase. On the other hand, thematic nominalizations may indeed require lengthy inclusions of all verbal valencies to avoid ambiguity. In (113),

> (113) We have [Mr]. Flamſteed [...] fully confirming the ſaid Parallax, both by **his Correction of [Dr]. Hook's Obſervations**, and by a greater Number of accurate Obſervations of his own (Whiston, 1715, p. 30; emphasis added).

the inclusion of both the subject and direct object is clearly a way to avoid ambiguity, which due to the high number of people that are mentioned in the paragraph is highly expectable. In other examples like (114),

> (114) To spectators situated somewhere on these parallels, the sun will be vertical, or in the zenith, twice in the course of **one revolution of the earth about the sun** (Bartlett, 1855, p. 32; emphasis added).

however, there seems to be no direct reason to include both valencies. It is well-known –now and at the time Bartlett published his book– that the earth revolves around the sun, therefore the inclusion of the second complement may be only the result of the author's wordiness. Looking at data analysis it might be noted that in nominalization NPs valencies are not as important as in verbal realizations. Nominalizations have other features that maximize their function in the text with very few grammar obligations.

Given that PPs modifying nominalizations are the most common way of encoding information about the process, special attention has been paid to their study. Results are shown in figure 49.

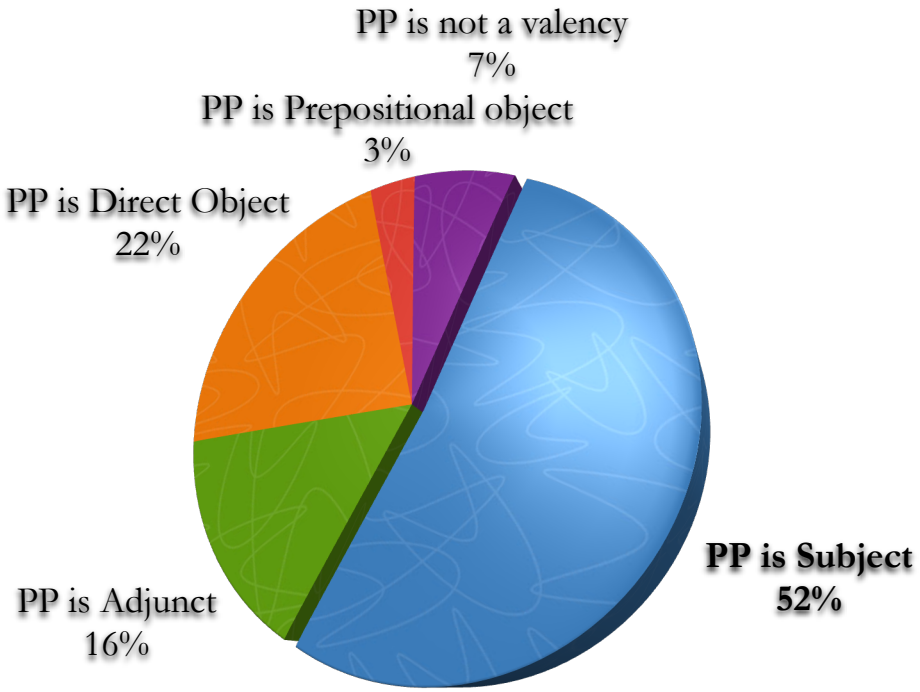


Figure 49: *Hypothetical function of postmodifying PP in nominalization NPs.*

It is striking that the percentage of postmodifying PPs that would function as direct objects in a verbal realization is less than half of those instances that would function as subjects. As far as word order is concerned, the duet *nominalized verb + direct object in Prepositional Phrase* parallels the canonical structure of verb and its complements. Also the postmodifying PP is usually paired with a possessive premodifying structure that provides an agent responsible for the action expressed in the nominalization, completing the SVO structure. This can be clearly seen in (27),

> (27) "This decision was fully borne out by Dr. Huggins's spectroscopic **observation of the disappearance behind the moon's limb of the small star Piscium, January 4, 1865** (Clerke, 1893, p. 324; emphasis added).

which could be rewritten as *On January 4, 1865, the small star Piscium disappeared behind the moon's limb. Dr. Huggins observed this with his spectroscope.*

Frequency data show, however, that no matter how appropriate this structure would be to respect word order patterns, this is not the most frequent complement. One reason for this is that not all verbs have the same number of valencies and, whereas, almost 100% of English verbs have a subject, the percentage of verbs displaying two or three valencies is quite inferior. This piece of evidence shows to which point it is statistically uneven to consider all verbal valencies at the same level. However, verb subcategorization was not included in this analysis because of the ambiguity it would entail. As Downing (1992, p. 72) remarked, verbs can have more than one type of subcategorization and by definition verb subcategorization is made according to the elements that appear in the Verb Phrase (Burton-Roberts, 1998, p. 80). Consequently, it would be ambiguous and unwise to give a verbal subcategorization of nominalizations without context.

Concerning other verbal complements (25%), prepositional complements (3%) are also present in postmodifying PPs. Surprisingly enough, there seems to be a correlation between the inclusion of this type of verbal complements and nominalization typology. Data show that that this type of complements are present when the head is either a term or a conditioned nominalization, that is, those nominalizations that are closer to nouns and that, in most cases have lost the semantics of process, usually associated with verbs. This is the case of example (115)

> (115) CHAPTER II. THE ASTRONOMICAL **OBJECTION TO RELIGION** (Whewell, 1858, p. 53; emphasis added).

where the nominalization is the keyword in the title. There are very few examples which can demonstrate better than a title all the properties of nominalizations. In this case, the nominalization summarizes and condenses all the information and presents it in a reified form that facilitates its cognitive comprehension by the reader. The inclusion of the prepositional complement *to religion* becomes very important here because it narrows down the topic of the text –the type of objection the text is going to deal with– and, at the same time links the two main actors of this process, *Astronomy* and *Religion*. It is interesting to notice again how, in spite of being the head of the NP, the role of the nominalization is that of linking the two main actors in the phrase.

Surprisingly enough, the frequency of postmodifying PPs that might function as adjuncts in a hypothetical verbal realization is relatively high[105]. As optional modifiers in the VP, it could be expected that PPs like *at the present period* in (117) would be left out in the nominal realization.

> (117) The **diminution at the present period** is about 52" in a century (Gummere, 1822, p. 235; emphasis added).

However, in this case again, the fact that adjuncts are included may indicate a difference in the construction of NPs with a nominalization as head. The importance of

[105] Place postmodifiers following a nominalization expressing a process of movement or event, as in (116)

> (116) "The two laſt articles being well underſtood, let us ſuppoſe the Earth to be projected in the ſtraight line igk, and the Moon to continue her **motion round the Earth**." (Ferguson, 1753:, p. 147; emphasis added)

have been considered adjuncts and not Prepositional Objects. Scholars show no unanimity on the complementation pattern of these verbs. The idea that these elements are not always present in the sentence lead to their categorization as possible adjuncts for this analysis.

semantics over other syntactic concerns may indeed be considered of paramount importance in the use of nominalizations.

Subject (52%) is the most frequently included valency in NPs containing a nominalization, partly because this is the only valency common to all verb subcategorizations. More than half of the postmodifying PPs are agency indicators. In (30),

> (30) The next announcement of the discovery of "Vulcan" was on the occasion of the total solar eclipse of July 29, 1878. [...] This time it was stated to have been seen at some distance south-west of the obscured sun [...] and **its simultaneous detection by two observers** —the late Professor James [C]. Watson, stationed at Rawlins (Wyoming Territory), and Professor Lewis Swift at Denver (Colorado)— was at first readily admitted (Clerke, 1893, p. 307; emphasis added).

the PP *by two observers* could be clearly rewritten as *two observers detected this fact*. The whole paragraph could be also rewritten as: *Vulcan was at the South-West of the sun during the total solar eclipse of July 29, 1878. Two observers detected this fact. The scientific community admitted and announced the event.* The preference of the nominal encoding, however, is not a matter of coincidence. In this case, the use of the nominalization was conditioned by the decision of situating this process at the beginning of a sentence to proceed later to validate it ("the detection [...] was admitted"). Thus, by reorganizing all the elements around the process in an NP, the process expressed in *detection* is no longer perceived as an ongoing process but as a reified entity that can be admitted.

The higher frequency of postmodifying subjects with respect to postmodifying direct objects shows to which degree nominalizations cannot be considered mere

transformations of verbal structures. Even if the semantics of the verb is to a higher or lesser extent maintained, nominalizations have their own organization, which may or may not coincide with the verbal realization. If nominalizations were mere transformations, a higher frequency of *possessive (Agent/Subject) + nominalization (process) + prepositional phrase (Thing/Direct Object)* would be expected. However the rate is higher than 2:1.

Given their presence both as pre- and postmodifiers and their high percentage of inclusion, an analysis of the presence or absence of agents together with its position in the phrase was extremely desirable. The results of this analysis are shown in figure 50, which shows that nominalizations indeed tend to occur without their subject being directly included. It may be noted that these data reflect direct agency inclusion into the pre- and postmodifying fields of the nominalization NP. The implication of this is that there may be other types of agency inclusion that have not been accounted for and fall within the *no subject transference* group. This is the case in (118),

> (118) These seem to have been the motives by which **Copernicus** was led to conceive the bold design of attributing motion to the earth [...]. This **attachment** indeed to the doctrines of uniform circular motion, which made him reject the excentric of Ptolemy, was merely a prejudice connected with the imperfect state of physical knowledge (Small, 1804, p. 84; emphasis added).

where the agent of *attachment* can be retrieved from the text some lines before. In other cases, the subject cannot be found in the text but it can be understood our shared knowledge of the world and does not need to be specified in the text, as in (119),

> (119) For the better **underſtanding** of this Matter, we will explain it by [Fig]. 13. Plate 6 (Gordon, 1726, p. 104; emphasis added).

where it is clear that the agent of *understanding* is *the reader of the book*.

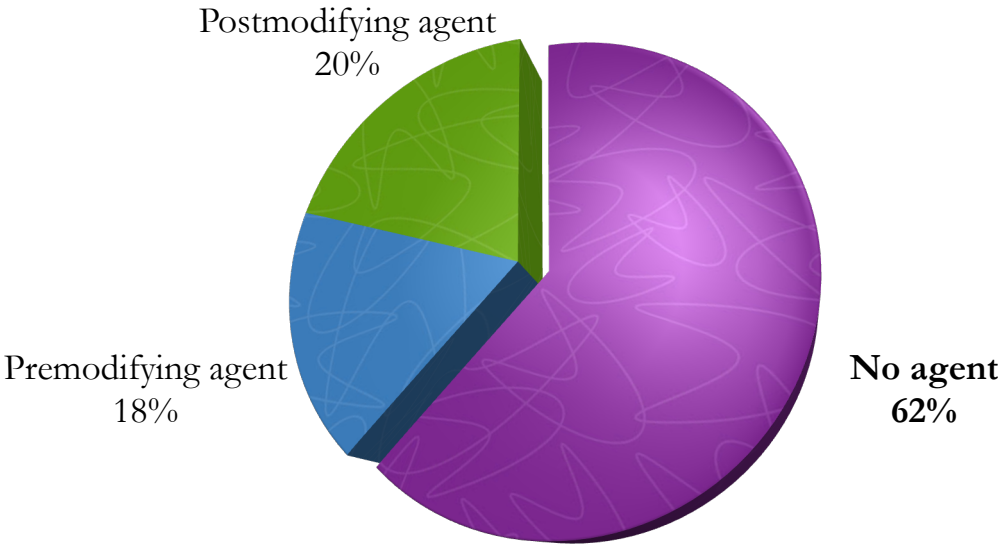


Figure 50: *Agent inclusion in nominalization NPs*.

Figure 50 shows that there is a slight tendency to accommodate agents in a postmodifying positions through PPs (18%) even if the premodifying field has more options for agent inclusion in terms of grammatical category. This fact points out that nominalizations have indeed more flexibility when it comes to organizing agents, processes and results in the phrase. The almost 1,700 instances of postmodifying agents all resemble the structure of (120),

> (120) Still minuter enquiry, however, detects yet smaller deviations again from this form and from these laws, of which we have a specimen in the slow **motion of the axis of the orbit** spoken of in [art]. 318 (Herschel, 1833, p. 206; emphasis added).

where the subject is contained in a PP introduced by the preposition *of*. On the contrary, the premodifying elements are very flexible and admit multiple subject transformations. In most cases the premodifying subject becomes a possessive and is expressed through a possessive determiner as in (121),

> (121) The periods of comets in **their revolutions** around the sun are equally various (Olmsted, 1841, p. 316; emphasis added).

where the referent has already been made explicit previously. If the referent needs to be specified, a Saxon genitive construction may be preferred as in (122),

> (122) [...] and in uſing the Hour Index for **the Sun's riſing and ſetting**, as before, the time of **Moon riſing and ſetting** will appear [...] (Charlton, 1735, p. 41; emphasis added).

Another option can be to include either a noun as in *Moon rising* in (122) or an adjective as in (123).

> (123) Unquestionably, the study of morals and the principles of **human action**, is at once the most important and dignified (Garland, 1838, p. 124; emphasis added).

These two last types are overtly less frequent in use, which indicates that the possessive element present in Saxon constructions and pronouns is a nuance in meaning that is favored and wanted. In some cases this could be refuted by pointing out that possessive pronouns usually fulfill a deictic function and are very useful devices for text cohesion. Even if this is undoubtedly true, examples like (122) show to which point the Saxon genitive and the nominal premodification are very similar in form and can co-occur in the same sentence. Despite this, nominal premodification is quite rare and

actually in (122) it may have been used for stylistic reasons, that is, to avoid the very same structure in the same sentence. After data analysis, it can be stated that compared with the encoding of processes into verbs, nominalizations have more possibilities for agent accommodation.

Regarding the inclusion of circumstances in nominalizations NPs, the 1,788 instances in which we can find circumstances may be considered as an indicator that phrases constructed around nominalizations do not necessarily share the properties of VPs. The elements appearing in VPs are normally dictated by their syntactic relationship with the main verb. If this may be also applicable to nominalization NPs, after having a look at figure 51, it may be stated that the parameters for determining what is a complement and what is a modifier in a VP do not necessarily get transferred to nominalization NPs. If nominalizations were only transformations of VPs, it would be expectable to find that adjuncts would be transferred to nominalization NPs only after some or all the verbal valencies had found their place in the NP. However, data analysis shows the inaccuracy of this claim and in some cases adjuncts are the only modifier found in nominalization NPs.

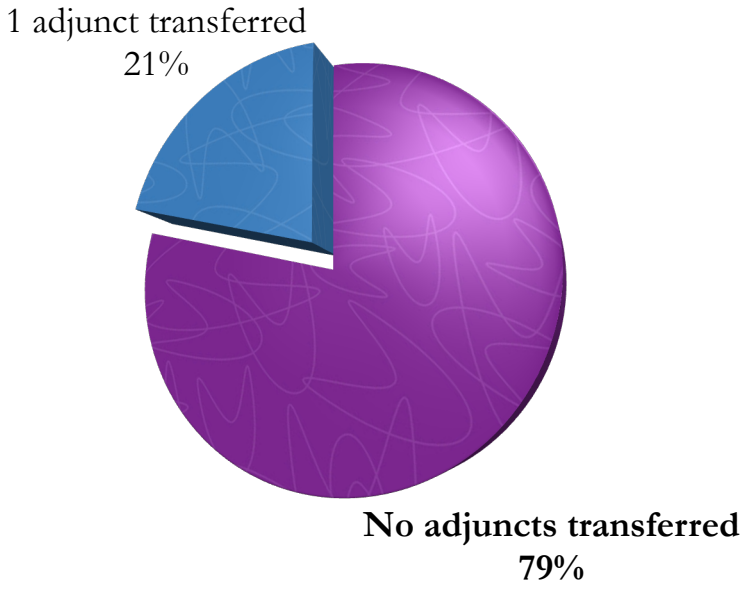


Figure 51: *Circumstance inclusion in nominalization NPs*.

Similarly, the existence of a postmodifying adjunct is not necessarily associated with the inclusion of any other verbal valency and, as a consequence, there are some examples in which the adjunct PP appears with another verbal valency, either a subject, as in (124)

> (124) Near this Conſtellation there are ſeveral unformed Stars, which in the year 1679. Mr. Edmund Hally, in memory of Charles II. King of Great Britain, &c. who was preſerved by **his Hiding in an Oak**, reduced them into a Conſtellation, and called it Robur Carolinum (Morden, 1702, p. 36; emphasis added).

or a direct object as in (125),

> (125) [...] 59 Seconds, will be the mean Motion for two Days, which ſtands againſt the 2d of January, and thus by the **continual Addition of 4 Minutes, 59 Seconds, 18 Thirds**, the mean Motion of the preceding Day, you will have the mean Motion of the ſucceeding Day [...] (Hodgson, 1749, p. 88; emphasis added).

while in others, the adjunct is the only modifier of the nominalization as in (126)

> (126) All the **appearances in the heaven**, both at land and ſea, are the ſame, as they would be, if the earth were a globe, which proves it to be of that ſhape [...] (Long, 1742, p. 63; emphasis added).

In the case of co-occurrence, there seems to be no restrictions regarding the position of the agent. Consequently in (124) we find a premodifying possessive agent, whereas in (127)

> (127) [...] in like manner, if the earth, which we conſider as always in the center of the ſphere of the heaven, has a rotation round its axis, ſuch a motion will cauſe to all the inhabitants of the earth an **apparent revolution of the ſphere of the heaven**, the contrary way, round the axis of the earth produced (Long, 1742, p. 71; emphasis added).

the agent is postponed and introduced by a preposition.

Perhaps one of the reasons for the high frequency of circumstance inclusion is connected to its adaptability to different realizations. Thus, circumstances can appear as APs in premodifying structures as in (125) and (127) or as PPs as in (124),(126), (128) and (129). Its co-occurrence with subjects is then mutually determined as examples (65), (66) and (69) show and it is probably related to the need of specification of the agent. Hence, if it can be condensed in a possessive construction, either as a pronoun (124) or a Saxon genitive as in (128)

> (128) And, therefore, the difference of **the Sun's attraction on the sides of the Earth under and opposite to him**, is much less than the difference of **the Moon's attraction on the sides of the Earth under and opposite to her**: [...] (Brewster, 1811, p. 263; emphasis added).

the circumstance can occupy a postmodifying position. In the case of (124) *his* refers to King Charles II and the substitution of the referent by a determiner has direct deictic and cohesive implications. In those cases in which the agent is made explicit through a PP, the adjunct is more likely to appear as an adjective before the nominalization (127), although it is possible to find instances of double postmodification, as in (129)

> (129) The unbounded view of nature, which I have laid open in my laſt letter, and the wonderful operations of the Deity in every part of this ſtupendous fabric, will not only ennoble the mind and ſtrengthen the underſtanding, [...] (Bonnycastle, 1786, p. 51; emphasis added).

All in all, the study of circumstance inclusion shows to which degree the distribution of processes, participants and circumstances is arranged in a different way in nominalization NPs. As far as modifiers are concerned, there seems to be no restrictions regarding what elements may pre- or postmodify the nominalization: definite articles and PPs are the most frequently used pre and post-modifiers, respectively. Nevertheless, the use of possessives differentiates nominalizations from other nouns. According to person they may have different functions as well: third-person possessives, which are the most frequent, have deictic functions and they tend to encode the agent of the process. First-person possessives are normally related to stylistic concerns, while second-person possessives are usually dictated by text-type. In the last

part of this section, attention was drawn to the inclusion of elements associated with the process expressed in the nominalization. Data showed that agents are the most frequent element and nominalizations, unlike verbs, provide multiple possibilities for their accommodation in the phrase, depending on what writers want to communicate. One surprising finding was the high frequency of circumstance inclusion. Circumstances are usually expressed in the VP as optional modifiers and therefore it could be expected that their accomodation in the nominalization NP would be less frequent. High circumstance inclusion rates show, however, that nominalizations are not mere verbal transformations. They are different ways of expressing processes and they have rules of their own, which are dictated on the writer's decision to focus on the process and allow the inclusion of certain information. Once all the variables have been applied to the total number of nominalizations, the next section will attempt at detecting and analyzing the main differences found across the different typologies.

4.2. Analysis of nominalizations according to typology

As explained in chapter two, a typology of nominalizations was created for this study following the belief that nominalizations are part of a continuum. In this process, they may lose morphosyntactic or semantic features of verbs and acquire those of nouns. This turning is not a mere transformation, as it responds to the writer's will to focus on the process and hinder agents, objects and circumstances. Additionally, nominalizations

have functions that neither verbs nor nouns display. In this study the following typologies are considered:

1. **Thematic nominalization**: it fulfills all the features described by functionalists (Banks, 2001, 2005): a process that is codified into a verb group (congruent codification) and functions as rheme can be turned into a nominalization (grammatical metaphor) in the theme of the sentences at the end of the paragraph. It also serves as a cohesive device, repeating and summarizing information and constitutes a perfect standpoint for the advancement of discourse. In (34)

> (34) But if the limbs be ſeperated, the object end is inclined to the quadrant, and muſt be **adjuſted** accordingly, and repeat the operation till the limbs coincide at both wires, and the **adjuſtment** is made (Vince, 1790, p. 15; emphasis added).

the advancement of discourse is achieved by turning the verb *adjusted* into the nominalization *adjustment* at the end of the sentence. With this twist, the process is reified, that is, acquires some of the semantics of nouns, which usually encode entities and things. Thus, its processing by the reader is simplified. Semantically, the proximity to verbal realizations is evident in this type of nominalization.

2. **Stylistic nominalization**: the choice of nominal over verbal realization in this type of nominalization is made to meet stylistic concerns. Stylistic complexity is considered a way of "guild codification", a code that only members of the community master and that distinguishes outsiders and novices from well-established members (Ventola, 1996). Stylistic nominalizations are thus complex markers of specialized discourse. In (37)

> (37) Her distance from the sun, like that of Ceres, is about 263 millions of miles, and she **performs her revolution** about it, nearly in the same time (Phillips, 1817, p. 68; emphasis added).

the author might have resorted to a simpler structure (*she revolves about it*). However, the choice of structural complexity in the shape of a light-verb construction containing a nominalization delimits the level of specialization of the writer, the audience and the text.

3. **Conditioned nominalization:** it is usually dictated by grammatical needs of the text. Extremely concise, they are usually the option writers use to condense several processes into one single sentence and omit unnecessary elements. In (39)

> (39) The same comet, also, came very near the earth; so that, had its quantity of matter been equal to that of the earth, it would, by its **attraction**, have caused the earth to revolve in an orbit so much larger than at present, [...] (Olmsted, 1841, p. 318; emphasis added).

the author wants to present the process as the agent causing another process (that the Earth revolves in an orbit). Since the agent, object and circumstances in which the process of *attraction* takes place can be retrieved from our shared knowledge of the world, the writer preferred to guide the reader's attention on the process itself by presenting it in the shape of a nominalization.

4. **Term nominalization**: this kind of nominalization is nearer to the semantic codification of entities as nouns. They can be labeled as "terms" as they are cognitive devices we create and use to study reality by establishing a set of

differences and frontiers (Lakoff, 1990). They are specially useful in scientific disciplines because they provide semantic traces of entity (reification) to both processes and entities (Banks, 2005b). Consequently they are functional guidelines that facilitate the organization of information in the mind of the reader. Thus, in (40)

> (40) But independent of theſe conſiderations, this rude ſyſtem was ſoon found incapable of ſtanding the teſt of **obſervation** and **experiment** (Bonnycastle, 1786, p. 59; emphasis added).

both *observation* and *experiment* are presented as fully reified processes. The focus is on the process itself, and information about agents, which can be retrieved from our shared knowledge of the world –we know that astronomers observe and experiment–, objects and circumstances, which are unknown, is omitted. By highlighting focus points the writer has provided, with the use of two nominalizations, hints on how to decode the text.

Figure 52 shows the general distribution of the 8,446 nominalizations in the four proposed categories. Stylistic (8%) with 704 occurrences and conditioned (11%) with 944 are the two groups with the lowest frequency. Thematic nominalizations (30%) are the second typology with the highest frequency of use.

Term nominalizations are ostensively the most populous group with 4,258 nominalizations, that is 50% of the total number of occurrences. The fact that the prototypical example of nominalizations –that is, those represented by thematic nominalizations– has such a considerable frequency difference regarding term nominalizations –the group with the highest frequency rate– may indicate that some of the most common features of nominalizations have indeed been neglected.

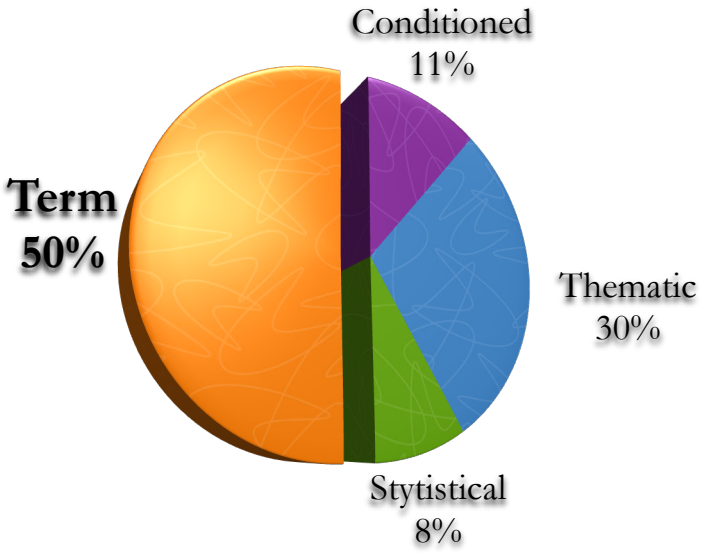


Figure 52: *General distribution of nominalizations according to typology.*

The general scholar conception of nominalizations has traditionally focused on the relationship between lengthy NPs governed by nominalizations and the VP, that is, the prototypical features of thematic nominalizations. However, they have failed to analyze other types of nominalization with less populated phrases and fulfilling peripheral functions which, however, show considerably higher frequency rates. This is perhaps one of the main claims of this study: that the referential role of those nominalizations in peripheral functions has been neglected and that they play a predominant role in text cohesion. Another premise that can be derived from the previous claim is that nominalizations are very flexible devices and, unlike verbs, they can express process meanings in a variety of forms and this can have multiple effects in texts and especially on the reader's interpretation of processes.

The evolution of all the different typologies, graphically represented in figure 53 below shows two simultaneous trends. On the one hand, the evolution of conditioned

nominalizations remains quite stable throughout the two centuries. On the other hand, lexical, stylistic and term nominalizations have a sharper increase over the two centuries. More specifically, the 8.66% increase of term nominalizations is the steepest of all, whereas conditioned nominalizations shows an increase of only 1.5%. However, it should still be pointed out that the evolution of all typologies is always increasing. This fact may help support the claim that nominalizations, in all their forms, showed a progressive steady increase in scientific texts from the beginning of the seventeenth century.

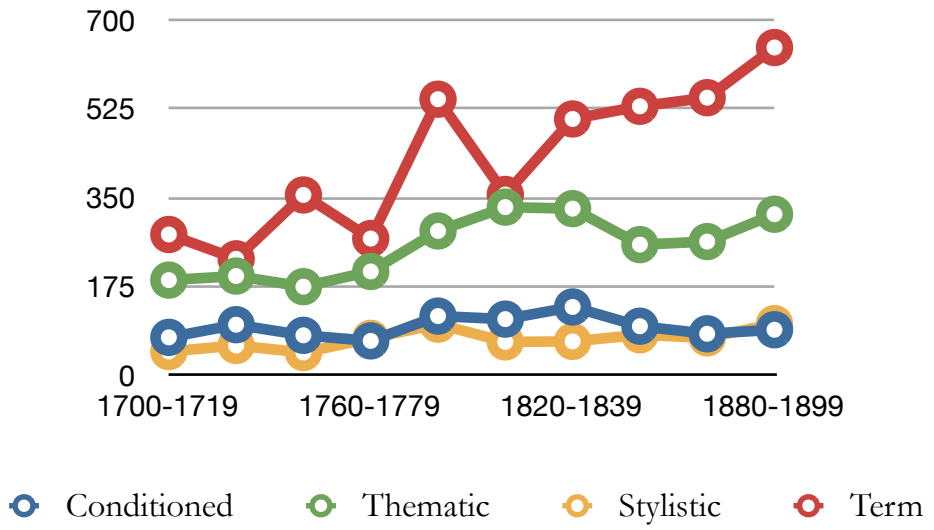


Figure 53: *General evolution of nominalizations according to typology.*

The explanation for the different evolutions of nominalizations may be related to the function of nominalizations in scientific texts. Despite nominalizations are now a widely used discourse marker for scientific texts and the reasons for this may have been obscured with time, it is true that the main features of thematic and term nominalizations originally fitted the needs of scientific texts: they provided texts with

coherence acting as lexical cohesive devices, packing and backgrounding information and, ultimately, making discourse advance. This is true for both lexical and term nominalizations in spite of their evident disimilarities concerning phrase structure and syntactic function. Considering the data in figure 53, it can be inferred that the increase of all these typologies may be related to these maximizing functionality premises.

There is no ground to argue that the parallel increase (7.8%) of stylistic nominalizations is also justified by the same functionality principles. In this case, the main defining feature of stylistic nominalizations is to provide a sense of belongingess to a specific discourse community for writers and, at the same time, define text-type, and consequently fields of expectation for readers. Considering that the first texts included in the corpus date from a period in which the register to be used in scientific discourse communities and even scientific communities themselves had not been standardized yet and that this regularization period partly took place in the timespan covered in the corpus, it is easily expectable that the increase rate of stylistic nominalizations may follow that of the discourse community it serves: a steadily progressive augmentation. It is also reasonable to believe that just as the scientific register and scientific communities became regularized, more distinctive linguistic features would pervade the language. In some cases, these distinctive linguistic features would clutter texts and ensure that texts would only be produced and understood by initiated community members. Other times, the inclusion of features typically associated with those topics and text-types would help readers categorize texts according to their text-type and, therefore, contribute to its correct processing. Regardless of possible motivations, it is clear that the increase in stylistic

nominalizations is related to extralinguistic features having to do with the establishment of a standardized scientific register.

Whereas lexical, stylistic and term nominalizations have increasing percentages of more than 5%, the timid 1.5% increase of conditioned nominalizations seems to be left out. The reason, however, seems more than logical. As it has been stated previously, nominalizations are well-known scientific discourse markers but their use has never been restricted to scientific texts or any other specialized language. They are present in all languages, in all registers and they are a sign of adult, abstract thought. Given their universality, a minimum rate of them is expected in language. Of all the typologies presented in this study, conditioned nominalizations are the ones whose functionality is less related to register. They are very useful because they help writers chain processes without compromising text structure. This function, however, is not more useful in scientific register than in any other register. Unlike thematic and term nominalizations, which maximize text potential and fit the purpose of scientific register, and stylistic nominalizations, which originated due to extralinguistic factors, conditioned nominalizations would be equally useful in other registers. This explains their low increase in the two centuries. It must be noted, however, that the increase, though small is still perceptible, which may be an effect of the general increase in the total number of nominalizations in the corpus.

Extralinguistic variables are not included in the study according to typology because their results were inconclusive. In the case of the geographical and gender variables, all typologies showed approximately the same values. Regarding geographical variation, all typologies have a slightly higher frequency in American texts –approximately 60% of the total with a maximum variation of 2% across typologies.

Then, the variable of sex of the author also produces very similar results: 58% of nominalizations belong to texts written by women and the maximum variation across typologies is also 2%. Concerning text-type, all typologies share a slightly higher frequency of nominalizations in formal texts rather than in learner's texts. Frequency percentages in formal texts are all around 53% (conditioned) and 57% (term). However, stylistic nominalizations show a slightly different behavior, as its frequency in formal texts amounts to 62% (438 occurrences). The reason for this small increment is clear: style is something more expectable in formal texts aimed at a learned audience that can appreciate intricateness of discourse. After data analysis, the differences among typologies seem not to have been dictated by extralinguistic variables. For this reason, the next sections will be purely devoted to linguistic concerns.

4.2.1. The morphosyntax of nominalizations according to typology

The first object of the study of this section is the morphology of the different typologies. In section 4.2.1.1 a brief descriptive account of suffix use is provided. Then special attention is paid to etymology of both roots and suffixes. In section 4.2.1.2 the topic of study is shifted to the syntactic functions fulfilled by nominalizations and by NPs governed by nominalizations in sentences.

4.2.1.1. The morphology of nominalizations according to typology

Suffix distribution does not show ample differences across typologies. In all four cases, *-ion* is exceedingly the most numerous suffix. 86% of stylistic, 83% of thematic and term nominalizations and a slightly inferior 73% of conditioned nominalizations were formed with *-ion*. Then, *-ing* is also fairly common in all typologies –13% of conditioned, 8% of thematic and 5% of term nominalizations– except for stylistic nominalizations, which only have 8 occurrences (1%) with this suffix. The explanation may be related to a possible restriction of the Germanic element in stylistic nominalizations. This etymological restriction applies not only to suffixes, but also to roots and hybrid combinations.

The distribution of the etymological origin of the root of nominalizations does not show vast differences according to typologies either. More than 85% of nominalizations have a root of Romance origin. There is no apparent reason to explain why conditioned nominalizations have a smaller percentage of Romance roots (89%). Similarly, given the small percentage difference between thematic (94%) and stylistic (95%) Romance nominalizations, a reasonable explanation seems unlikely to be found. The case of stylistic nominalizations, however, presents a slight modification. The percentage of Romance nominalizations in the stylistic type (99%) is significantly higher. There are only 4 instances (*ending, meaning, beginning and understanding*)–out of 704 the total number of nominalizations in this group–of nominalizations with a root of Germanic origin. This fact is even more surprising considering the prototypical structure of stylistic nominalizations: a light-verb construction where the nominalization is the

Object of a semantically-emptied verb which, in most cases is of Germanic origin. Thus, the low percentage of Germanic stylistic nominalizations may be linked to an apparent incompatibility between verbs and nominalizations of the same origin

The etymological analysis shown in figure 54, of roots and suffixes is not very revealing. As expected, combinations of roots and suffixes with the same etymology prevail and hybrids have less frequency of use. Hybrids repeat the same pattern in all typologies. Nominalizations with a Romance root and a Germanic suffix, though scarce, are the most repeated type of hybrid, whereas the opposite is so scant in number that there are only three instances, namely *amazement*, *acknowledgement* and *wonderment*. The three are term nominalizations but rather than to restrictions according to typology, probably as a result of corpus size.

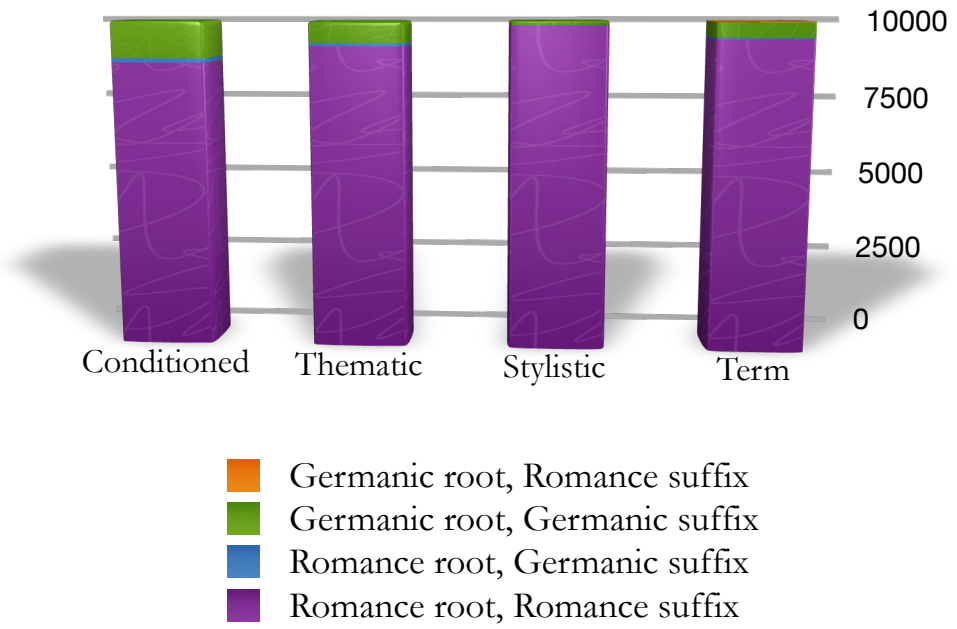


Figure 54: *Distribution of etymology of roots and suffixes according to typology (NF 10,000).*

Nominalizations with Romance root and suffix are, as expected, the most common combination in all typologies. Conditioned (87%) and stylistic (99%) are the

groups with the lowest and highest frequency of purely Romance nominalizations, respectively. Concerning purely Germanic nominalizations, conditioned nominalizations (11%) are undoubtedly the most populous group. The rest of the typologies all have ostensively smaller frequencies of purely Germanic nominalizations. Stylistic nominalizations (114 occurrences) is the group with the lowest frequency.

The awkwardness of stylistic nominalizations is consistent as far as hybrids are concerned. There are no occurrences with a Germanic root and a Romance suffix in this typology and only fourteen occurrences with a Romance root and a Germanic suffix. These data may suggest that stylistic nominalizations have some exclusion rules regarding etymology. This type of nominalization is mostly formed with Romance roots and typically combined with Romance suffixes. However, instead of aiming at general hypotheses about structure, in this study one plausible explanation may be related to the topic of texts (astronomy) and to the fact that stylistic nominalizations are markers of specialized discourse. The most repeated stylistic nominalizations are *motion* (in the construction *have a motion* or *perform a motion*) and *observation* (in *make an observation*). The case of *motion* seems especially indicative as the choice of verb (*have* or *perform*) is linked to the information about the process that is included in the nominalization phrase. Thus when information about circumstances in which the process took place is to be included, then the most frequent collocation is *to have*. In (130)

> (130) [...] he [the Sun] ſeems to **have an annual motion** in the heavens, and to riſe and ſet continually in different parts of them (Bonnycastle, 1786, p. 29; emphasis added).

the verbal time adjunct is included as an adjective (annual) premodifying the nominalization and it, together with the nominalization, selects the semantically-emptied verb *to have* to form the light-verb construction. On the other side, the verb *perform* is usually paired with a possessive premodifying the nominalization. This is the structure of (131)

> (131) [...] it will appear from the following Table, that the Sun [...] muſt be the Body to which the other Planets gravitate, or the Center about which they **perform their Motions** [...] (Gordon, 1726, p. 114; emphasis added).

where the light-verb construction includes a reduplication of the agent (*other Planets*/ *their*).

These two (*move* and *observe*) are the two most common processes in astronomy: celestial objects move; astronomers observe them (both the planets and their motions)[106]. Given their frequency in texts, it seems thus expectable that stylistic variations have arisen. Their function is double. As lexical cohesive devices they constitute a synonymous expression of highly frequent words. Besides, at the sociolinguistc level, they mark addresser/addressee boundaries and serve as guild codification. As a result, in this study, the peculiar etymological distribution of stylistic nominalizations is connected with the requirements of the topic of the text.

With the particular exception of stylistic nominalizations that seem to have some etymological restrictions and more refined ways of arranging its components, morphology does not seem to be a very indicative variable for typological

106 The definitions of astronomy acknwledge that observing and dealing with the motions of the planets is the main task of the discipline. Thus, the *OED* includes the reified nominalization "motion" in the definition: "Astronomy: the science or art dealing with the positions and motions of planets and stars and their effect on natural phenomena and human affairs (one of the subjects of the quadrivium; now hist.)" (*OED online*)

differentiation. Suffix choice and etymology show similar results across typologies. The next variable, syntax, has given very different results according to typology.

4.2.1.2. The syntax of nominalizations according to typology

Figure 55 shows the distribution of the four categories in the four most common syntactic functions, namely, subject, direct object and adjunct. Compared with other parameters showing very little variation among categories, the syntactic function is another focal point for typology discrimination. Nevertheless, all typologies have a high frequency of nominalizations functioning as adjuncts or peripheral modifiers within sentences: 67% of conditioned and 68% of term nominalizations belong to this group. adjunct is also the most populous function (43%) within the group of thematic nominalizations. Only in stylistic nominalizations it can be found that direct objects (56%) have a higher frequency than that of adjunct (13%).

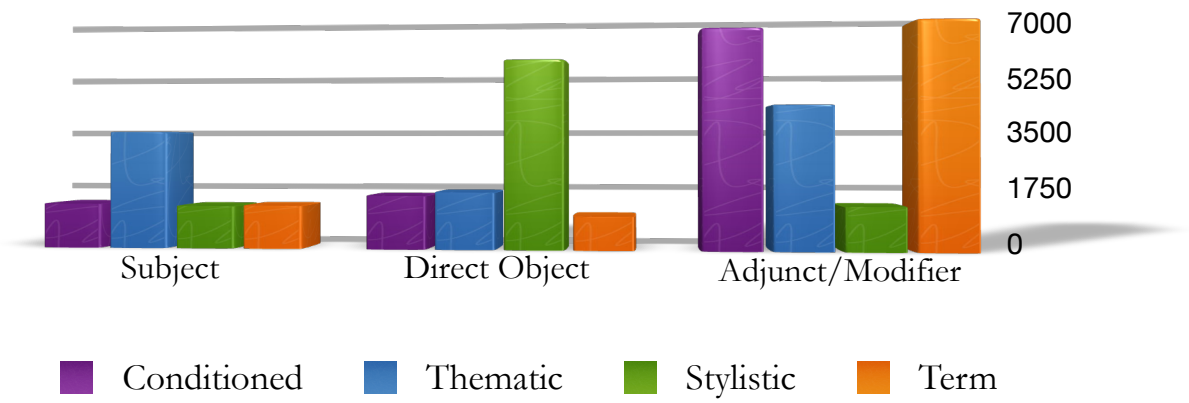


Figure 55: *Distribution of syntactic function according to typology (NF 10,000).*

It seems appropriate to highlight the importance of this finding: that contrary to all the traditional bibliography centered on examples of nominalizations functioning as subjects and subject complements, frequency data backs up the idea that nominalizations tend to fulfill peripheral roles in the sentence. The methodology used for this study has made possible this new perspective. One of the main benefits of corpus linguistics is that it enables the extraction of more reliable data backed up by frequency numbers.

This claim, of course, does not minimize the validity of the findings about nominalizations that have been made so far. On the contrary, it builds on them to complement our understanding of nominalizations. Ravelli (1988) cited conciseness as one defining feature of nominalizations. In the case of nominalizations functioning as adjuncts this conciseness may be due to the fact that nominalizations are focalizers of information. Thanks to their flexibility concerning phrase structure, they manage to focus attention on the process meaning and avoid referring to agents, results or other circumstances. That way, they provide the reader with hints on how to process information and decode the text and avoid ambiguities and distractions. In the case of adjuncts and modifiers, this focalizing of information is somehow more consistent with their peripheral role in the sentence. In (40)

> (40) But independent of theſe conſiderations, this rude ſyſtem was ſoon found incapable of ſtanding the teſt of **obſervation** and **experiment** (Bonnycastle, 1786, p. 59).

both *observation* and *experiment* appear in a peripheral role and the specification of the agent and the object seem irrelevant for two reasons. On the one hand, they can somehow be identified following our shared knowledge of the world (it may be inferred

that scientists/astronomers would observe the celestial objects and experiment their findings/hypothesis). The important focus of information here lies on the expression of the process, not on their agents, objects or other circumstances. Their inclusion would hide the focus of attention and complicate the reader's understanding.

Concerning the application of this focalizing feature, the high percentages of nominalizations in adjunct positions of conditioned (67%) and term nominalizations (68%) seems logical. Unlike thematic nominalizations, which are a rearrangement of a verb into a nominal with different purposes (establishing a point of reference for the advancement of discourse, packing information), and stylistic nominalizations, which are text-type markers, conditioned and term nominalizations make great use of their focusing features and develop it in peripheral roles within the sentence. In the case of conditioned nominalizations, this focusing strategy is dictated by a need to reduce useless verbal valencies and to avoid grammatical complexity. In (132)

> (132) Collecting confidence from these circumstances, he announced his discovery as the result of **observation** and **calculation** combined, and entitled to as much confidence as any other consequence of an established physical law (Olmsted, 1841, p. 327).

the agents of the nominalizations can, as in previous examples, be easily inferred from common knowledge of the world, which renders them optional. Unlike (40), the grammatical structure is more complex. In this case there are five processes directly mentioned in the sentence: observing the celestial objects, calculating the results of the study, discovering new findings, collecting confidence and announcing the findings. In the sentence the chronological order is altered, almost reversed. However, since all processes share the same agent –they are referring to the famous astronomer Halley who

appears in the previous paragraph–, they can be rearranged in the same sentence. Of all processes, only *announced* is a conjugated verb governing a sentence. *Collecting* appears as a non-finite verb and the rest are nominalizations (*discovery*, *observation* and *calculation*.

As far as term nominalizations are concerned, the focusing features join in the fully reified form of the nominalization and all its cognitive implications. Following the strict English SVO pattern, term nominalizations functioning as adjuncts appear at the end of the sentence. In this position, the nominalization has lost its semantic verbal properties, as well as all verbal valencies or modifiers, and is presented as a static fact. As Banks (2011) pointed out, its conceptualization changes and they become perceived as an object functioning as a circumstance in another process, as in (133)

> (133) At this time Dr. Archibald Geikie took up the question and **went into the consideration** of the subject in a most thorough manner; and it is mainly through the instrumentality of his writings on the matter [...] that the method under **consideration** has gained such wide-spread acceptance among geologists (Croll, 1889, p. 41; emphasis added).

where *consideration* appears twice. At the begining (*Dr. Archibald Geikie went into consideration*) *consideration* is presented as a stylistic nominalization inserted in a light-verb construction. (*Dr. Archibald Geikie*) and the object (*the subject*) are made explicit because the author is narrating a series of events. The second time *consideration* appears in the sentence, it has lost all its modifiers and verbal semantic properties. It is now a term nominalization functioning as a modifier. The shift of focus, however, is evident: the author is focusing on the reified features of the noun. He wanted to point out that the process is unquestionable, a static fact, not a mutating

process. In order to do so, he inserted a reified nominalization as modifier in the subject of the relative sentence. The process has become more reified in the last part of the sentence and it is also rearranged to constitute a common ground for the advancement of the discourse. It may be noticed, however, that the subject and main focus of this relative sentence (*that the method under consideration has gained such wide-spread acceptance among geologists*) is *the method,* not *the consideration*. The referential role of this term nominalization is thus maximized. Including this term nominalization as a modifier, the author is subtly making a shift of focus but still keeping all the referential implicatures. This is the main difference between nominalizations functioning as adjuncts and those in subject roles.

The syntactic function is in many cases correlated with the use of semantically-emptied verbs. The distribution of such verbs can be seen in figure 56.

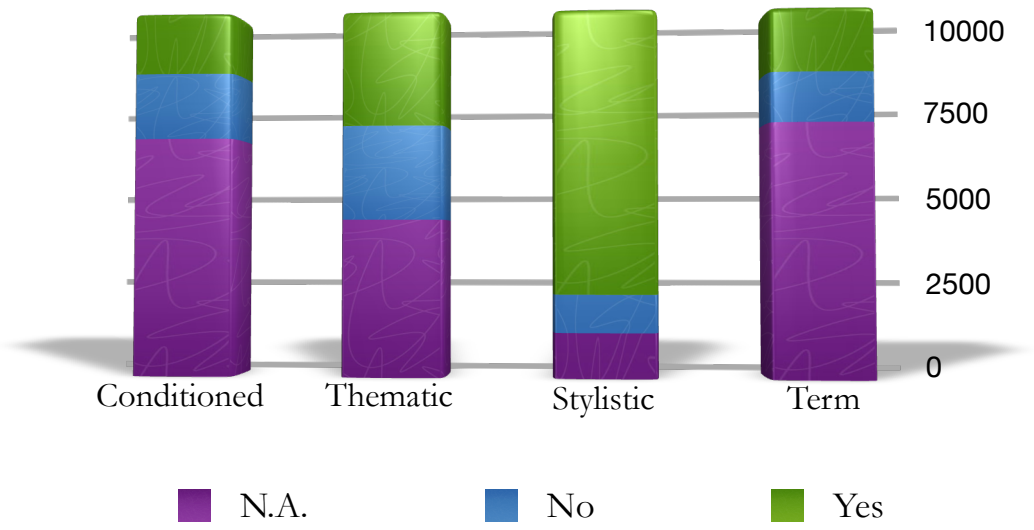


Figure 56: *Distribution of semantically-emptied verbs according to typology (NF 10,000).*

The high frequency of semantically-emptied verbs in stylistic (77%) and thematic (31%) nominalizations is motivated by the existence of two types of such verbs which can combine with two types of nominalizations. On the one hand, stylistic nominalizations are in many cases part of light-verb constructions, where the verb has lost most of its verbal semantics, which are taken over by the nominalization, as in (37)

> (37) Her distance from the sun, like that of Ceres, is about 263 millions of miles, and she **performs her revolution** about it, nearly in the same time (Phillips, 1817, p. 68).

In (37) the asteroid *Pallas*, mentioned in the title of the section and substituted in (37) by *she*, is said to revolve around the Sun. Apart from a subtle reificative touch, there is no functional or formal reason in the text that explains the preference for a nominalization other than presenting a stylistic alternative for the expression of the process. Whatever the reason may be, the high frequency of semantically-emptied verbs in stylistic nominalizations (77%) is undoubtedly a consequence of light-verb constructions where the nominalization takes on most of the semantics of the verb.

However, this is not necessarily always the case. Sometimes, the semantics of the verb is not taken over any other element in the sentence. As Halliday (2004) showed, the scientific register is progressively adopting the tendency to emphasize the relational aspect of verbs. Standard scientific English sentences are usually formed by two heavy nominalization phrases joined by a semantically-emptied verb that establishes the type of relation between the processes expressed in the nominalization NPs. In (93)

> (134) The **solution** of the problem of longitude **consists**, therefore, in finding the difference of the local times which exist

> simultaneously on the first and required meridians (Bartlett, 1855, p. 21; emphasis added).

the verb *consists* separates the sentence in two lengthy parts –*the solution of the problem of longitude* and f*inding the difference of the local times which exist simultaneously on the first and required meridians*–, while it also establishes the consecutive relationship between them. For the reader, the verb helps conceptualize the temporality in which both processes should be understood (first we find the difference, and then we will find the solution)

In some way, it is not true that the verb loses its semantics but rather that it acquires a new meaning –that of expressing a relation– while its former semantics is expressed by the nominalizations functioning as subjects or objects. This is the main reason why thematic nominalizations have such a considerable percentage (31%) of semantically-emptied verbs. In this typology, the number of verbs retaining its semantic properties is still high (26%) but this is most probably the result of the fact that the readjustment of the semantic value of verbs in scientific English is a trend that started 400 or 500 years ago (Halliday, 2004, p. 174). By the time the texts in the *CC* were written, this was still an ongoing process, as frequency data can here corroborate.

The semantics of the verb, however, do not seem to be specially relevant in the case of conditioned or term nominalizations. In 65% and 69% of them finding out the main verb of the sentence in which they appear is even difficult because they appear in embedded structures and usually far from it. This type of nominalizations perform, as we have just seen, very frequently peripheral functions. In those cases, whether the main verb of the sentence has retained its semantic value, transferred it to another

element or altered it seems less relevant. In these cases, their task is closely related to serve as a reference to other parts of the text.

This section has cast some light on how the role of verbs and syntax are powerful variables for typology differentiation. One of the most important findings may be the high frequency of term nominalizations in adjunct positions, which points out at the deictical role that term nominalizations play as focalizers of information. On the other hand, the role of verbs can also help recognize stylistical and thematic nominalizations. The latter are easily recognized because they combine a subject position with a semantically-emptied verb that relates it with the second part of the sentence. In the case of stylistical nominalizations, the nominalization usually functions as the direct object in a light-verb construction.

4.2.2. The morphosyntax of nominalization NPs according to typology

In this section, the structure of the nominalization NPs in the different typologies is the main object of study. Consequently section 4.2.2.1 is related to the description of the pre- and postmodifying field. The inclusion of a possessive structure is taken into special consideration, given the considerable differences found across typologies. In the last section (4.2.2.2), the emphasis lies on typological differences in agency and circumstance inclusion. The notion of verbal valencies is not so evident here as the main objective in this section is to study the ways different typologies have to include information about the process in their NPs.

4.2.2.1. The structure of nominalization NPs according to typology

The structure of nominalization NPs revolves around two poles: the premodifying and the postmodifying field. In both cases, differences among typologies are mainly dictated by their similarity to the semantics of the VP encoding. As a result, thematic and stylistic nominalizations are most likely characterized by lengthy modifiers and, most importantly, these modifiers tend to encode information about the process expressed in the nominalization. This information may be related to agents, objects, participants or circumstances around the process and, in case of a verbal realization, they would be transformed into verbal valencies or adjuncts within the VP. On the other hand, conditioned and term nominalizations, which are closer to the semantics of a noun, have less and shorter modifiers. In the case of conditioned nominalizations, scarcity is dictated by the need to adapt to a particular grammar context. These nominalizations are important points of reference in texts and this focusing strategy sometimes results in the omission of those modifiers that cannot fit in the context. Concerning term nominalizations, modifiers do not normally refer to the process but rather to the nominalization as a fully reified entity.

In the case of premodification, all typologies have in common a widespread use of determiners as the only element in the field (41% of stylistic and term, 62% of thematic and 69% of conditioned). More significant differences are to be found in the use of adjectives and possessive constructions because these two elements normally encode the agent or a circumstance in the process. AP inclusion represents 23% of conditioned, 31% of lexical, 32% of term and 40% of stylistic nominalizations. The low AP

frequency in conditioned nominalizations is, as discussed, a result of grammar constraints. In the case of thematic nominalizations, APs may include the agent of the process, as in (135)

> (135) The question of the sun's temperature, though connected with that of the **solar radiation**, is yet distinct, and presents still greater difficulties (Young, 1880, p. 93; emphasis added).

where the Sun is said to radiate. The percentage of AP inclusion is low (31%) because agents are most frequently included in possessive constructions or in postmodifying PPs. Nevertheless, in most cases, APs indicate a circumstance implied in the process. Such is the case in (95)

> (136) One would have thought that its **assiduous cultivation** by such men as Newton, Halley, Delambre, Lagrange and Laplace, might have rescued it from a contempt like this (Garland, 1838, p. 123; emphasis added).

where the adjective expresses a circumstance (assiduity) about how the *cultivation* took place. This is indeed a rare example of disruption of the SVO pattern in which the nominalization is modified by the object of the process (*its*, referring to astronomy, mentioned in the previous sentence), a circumstance (*assiduous*) and the agent in a postmodifying position (*by such men as Newton, Halley, Delambre, Lagrange and Laplace*). Most of the 41% of stylistic nominalizations including an AP also feature a circumstance about the process. However, in term nominalizations, which also have 41% of AP inclusion, the tendency is that APs are attributes to the reified nominalization. Thus, in (137)

> (137) For, of the arguments [...] have never [...] been put forth in that **religious bearing** which seems to belong to them; [...] (Whewell, 1868, p. 42; emphasis added).

the adjective *religious* is an attribute of a reified state (*bearing* is considered a state, not a process, and the attribute *religious* is attributed to it).

Possessive constructions are a useful indicator of typological variance. Unlike APs, which can encode agents, circumstances or attributes of reified nominalizations, possessive constructions are less ambiguous. The distribution of possessive constructions in the different typologies is shown in figure 57.

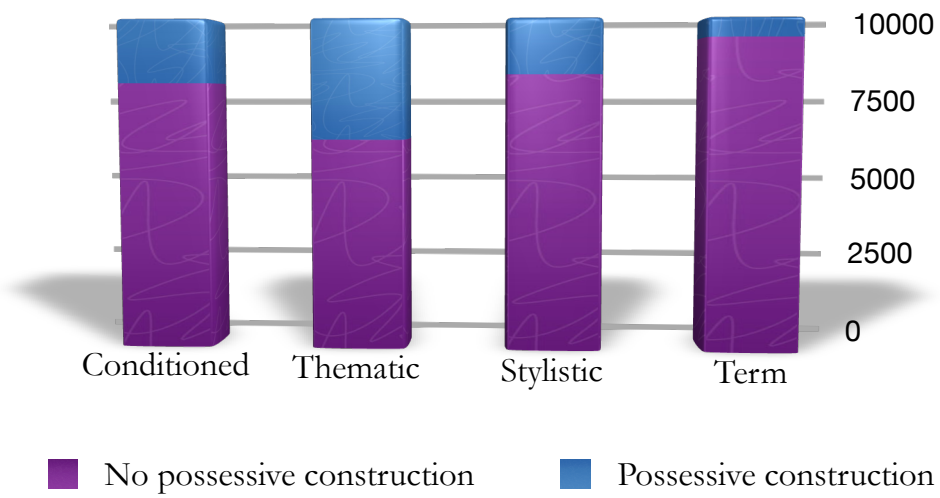


Figure 57: *Distribution of possessive construction according to typology (NF 10,000).*

As discussed in previous sections[107], the role of possessives changes completely according to person, number, and gender. Third-person possessives are lexical cohesive devices; their referent can be usually retrieved from the immediately preceding context. Given the special function that nominalizations play in scientific register as cohesive devices, conditioned, thematic and stylistic typologies have a greater frequency of third-

[107] See section 4.1.3.1.

person possessives. The sum of *his*, *her*, *its* and their is very high in all this three typologies: it amounts to 70% of conditioned, 69% of thematic and it reaches 91% of stylistic nominalizations. Additionally, *its* and *their* are the most repeated possessives in the three groups. The role of possessives may also be due to stylistic reasons, as in (138),

> (138) The Ecliptic is that great Circle in whoſe Plane the Earth performs **its annual Motion** round the Sun (Fuller, 1732, p. 6; emphasis added).

where the referent of *its (the Earth)* is found in the same sentence. The whole sentence could be rewritten as *the Earth moves* and the rewording seems to be a consequence of textual pompousness.

In the case of term nominalizations, the frequency of third-person possessive premodification is considerably lower (57%) as possessive determiners are replaced by Saxon genitives and mostly by *of*-constructions (22%, the highest frequency across typologies). Third-person possessive determiners usually establish or recall a syntactic relationship with its head, a feature that is unparalleled in most *of*-constructions. This is the case in (139),

> (139) The zenith distance is the **complement of the altitude** (Loomis, 1868, p. 12; emphasis added).

where there is no possible verbal rewriting of *complement* and *altitude.* As a result of this lack of verbal semantic properties, *of-constructions* are more useful in term nominalizations because, among all the typologies, this is the one that is semantically further from verbs.

In the rest of persons, possessives do not have many differences of either use or frequency across typologies. Second-person possessives, which are normally dictated by text-type show very similar frequencies of use, ranging from 1% of conditioned to 3% of term nominalizations. On the other side, first-person singular possessives also have very low frequencies (1%) in all typologies, whereas the possessive *our* represents around 5% in every typology.

Concerning postmodification, shown in figure 58, there is significant variance among typologies that ranges from 39% in term nominalizations to 68% in the case of conditioned nominalizations. Besides, the distribution of different types of postmodification varies according to typology even though in all typologies, PPs are the most repeated option (a minimum of 70% of the total)

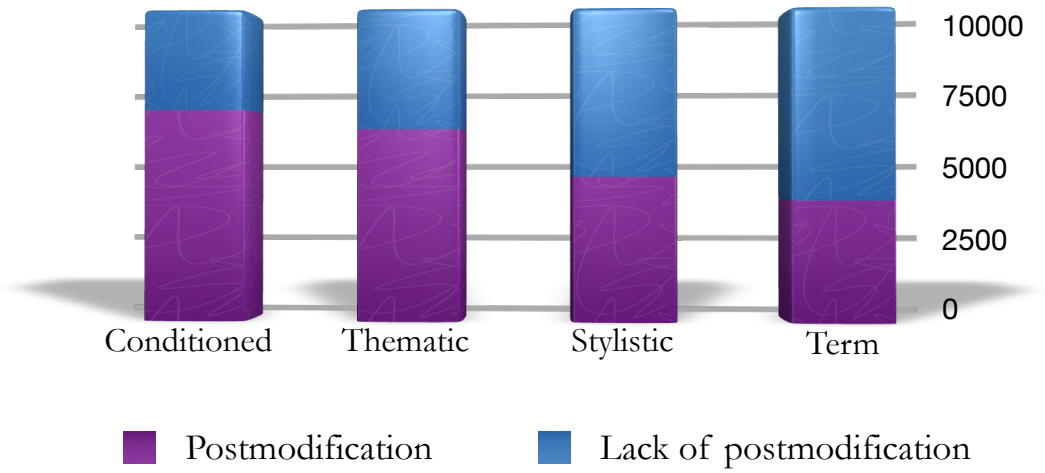


Figure 58: *Distribution of postmodification according to typology (NF 10,000).*

Postmodification, however, should not be measured as a completely independent variable. The high percentage of premodification of conditioned nominalizations (68%) is usually correlated with a low percentage of premodification (19%). Conditioned

nominalizations are usually dictated by grammatical needs of the text and therefore, it is logical that restrictions concerning double modifiers may arise. As in other variables related to the structure of nominalization NPs low frequencies are motivated by formal restrictions concerning the grammatical context. This is also why this type of nominalizations have the lowest range of available structures: 91% of postmodifiers are PPs and a low 9% are relative clauses. After all, behind the choice of the conditioned typology, there is a decision to focus on the meaning of the process. In some cases, this provokes an omission of modifiers that would convolute the text and deviate the attention from the process.

The second biggest group with a high postmodification rate is thematic nominalizations (62%). This typology has high frequency in both pre- and postmodification (62%) because in most cases these modifiers accommodate information about agents, objects, attributes and circumstances surrounding the process. Given their proximity to verbal realizations, thematic nominalizations have a greater tendency for double modification and that is reflected in the high percentage in figure 58. As far as postmodificacion options are concerned, 96% of all instances are PPs and the remaining occurrences display a relative clause. Among all typologies, this is the one with less postmodification variety. PP postmodification –usually starting with *of*– usually indicates the inclusion of former verbal valencies expressing agents or objects of the process. Thus, in (140)

> (140) [...] by diſcovering the cauſe of **their riſing on the ſide of the Earth oppoſite to the Moon** (Ferguson, 1756, p. 146; emphasis added).

rising is accompanied by a possessive indicating the agent and a postmodifier containing a spatial circumstance. The similarity with the verbal realization is enhanced by the order of the elements that respects the canonical SVO pattern.

The other most common type of postmodification, relative clauses, is, on the contrary, an indicator of reification. Relative clauses normally indicate that the nominalization is closer to the semantics of the noun, as the referent of the relative pronoun tends to be an entity or a highly reified process, that is, a nominalization. Cognitively, it is easier to refer to entities than it is to processes (Banks 2005). Hence, even term nominalizations show very low postmodification rates (37%), they have the highest percentage of relative clauses (20%). In (137)

> (137) For, of the arguments [...] have never [...] been put forth in that religious **bearing which seems to belong to them** [...].” (Whewell, 1868, p. 42; emphasis added).

it is clear that the relative clause refers to the reified meaning of *bearing* as a state, not as a process. This is verified by the attributive function of the premodifying adjective (nothing is born religiously, but rather *religious* becomes an attribute of a reified state)

Lack of postmodification is also related to reification processes. This is why term nominalizations have the biggest percentage (64%) of non-postmodified nominalizations. Considering that modifiers in nominalization NPs tend to include information about agents, objects and circumstances of the process, it is highly expectable to believe that term nominalizations, which barely retain the semantics of the process, may be the typology with a lower postmodification rate.

4.2.2.2. The functions in nominalization NPs according to typology

The distribution of the different typologies concerning agent inclusion, shown in figure 59 is quite irregular. From the scarce 17% of term nominalizations containing an agent either in the premodifying or postmodifying field to the overwhelming 70% of thematic nominalizations, agent inclusion seems to be a focal point for typology discrimination.

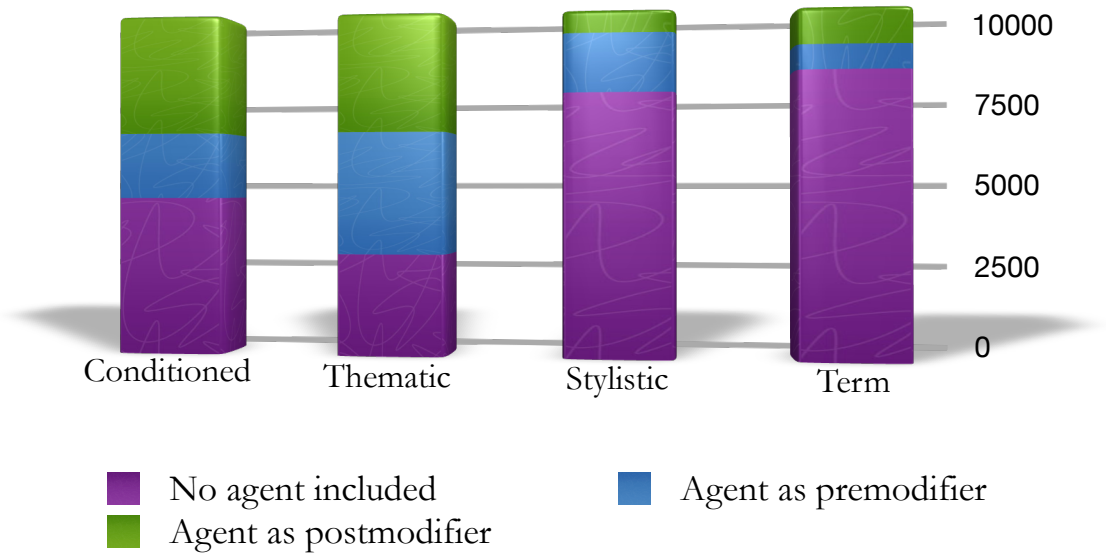


Figure 59: *Distribution of agent inclusion according to typology (NF 10,000).*

The main distinguishing criterion for agency inclusion seems to be the maintenance of verbal properties: thematic nominalizations still retain a great deal of semantic verbal properties and one of them is indeed the reincorporation of a subject valency in their phrase. That is why 70% of them include an agent. However, thist seems to be more important than the actual form in which it is encoded. 36% of thematic nominalizations are premodified by an AP or, most commonly, a possessive construction as in (141)

> (141) In the year 1742, about 10 years after **Mr. Hadley's invention** was publiſhed, a paper in Sir Iſaac Newton's own hand writing was found among Dr. Halley's papers after the Doctor's death, containing a figure and deſcription of an inſtrument Mr. (Vince, 1790, p. 6; emphasis added).

or are followed by a possessive construction in the shape of a PP governed by *of* as in (142)

> (142) We may know in what eſteem this **invention of Bayer's** ought to be held, when we obſerve, that all have publiſhed large or correct figures of the conſtellations ſince the time of their inventor, have continued them (Hill, 1754, p. 3; emphasis added).

Agentless constructions such as (143)

> (143) ASTRONOMY, which is deſervedly eſteemed the moſt noble and exalted branch of human literature, regards the various phœnomena of thoſe heavenly bodies, which the **invention** of curious inſtruments hath brought within our obſervation, from the ſurface of the terreſtrial globe (Adams, 1777, p. 1; emphasis added).

are less common in thematic nominalizations. In (102) the reason for the lack of agent is that they (the inventors of curious instruments) may not even be known for the author. Despite they are less numerous than those nominalizations with an explicit agent, sentences like (143) have received much attention, especially by critical discourse analysts (Billig, 2008; van Dijk, 1988, 2006; Fairclough, 1992; Fowler, 1991; Wodak & Meyer, 2001). Some scholars have considered this agentless nominalization a mystification device to conceal information. Thematic nominalizations are the closest typology to verbal realization and therefore have attracted most scholar’s attention with different results: Lemke (1995) pointed out at an overt information gap performed when

the author chooses to replace a verb with a nominalization. Frequency data obtained in this study –30% of agentless lexical nominalizations– seem to downsize the application of this claim[108].

Van Dijk (2008) acknowledged the effects of agentless constructions but he also noted that behind a nominalization choice other parameters like economy or lack of information may be involved. In (143) it is clear that a combination of these two may have been the reason for subject omission. Being the first paragraph in the book, it is reasonable that the focus of interest should lie on the definition of astronomy and not on the specification of all the inventors of astronomical instruments. In some sense, the absence of the agent responds to a higher need for information structuring After all, the understanding of utterances is not automatic and the reader needs "instructions" to perform their process of understanding. In this case, utterance decoding restrictions come in the shape of agentless nominalizations that force the reader to focus on the main topic of the paragraph and avoid extra processing costs by eliminating unnecessary elements.

Stylistic nominalizations seem to prefer for agentless constructions, as more than two-thirds (77%) of them do not include an agent within their phrase. This claim however cannot be accepted without a broader consideration of this special type of nominalization, mostly made up by light-verb constructions that include a nominalization as an object within the construction. This claim turns the data in figure 59 into a sort of mirage that does not represent those nominalizations whose agent is the grammatical subject of the light-verb construction. Such is the case in (103)

[108] This, of course, does not minimize the importance of fidings of CDA scholars. It is true that agentless constructions can have important mystificative implications but, looking at frequency data, this does not seem to be their main/only property.

> (142) A second proof of the earth's rotation is derived from the motion of falling bodies. If **the earth had no rotation** upon an axis, a heavy body let fall from any elevation would descend in the direction of a vertical line (Loomis, 1868, p. 33; emphasis added).

where *rotation* would be computed as lacking an agent because it does not appear in either the pre- or postmodifying fields surrounding it. However, our knowledge of the world tells us that the agent for this nominalization coincides with the grammatical subject in the sentence. In this particular case, the use of stylistic nominalizations seems to be chosen with the aim of avoiding the same structure with the same nominalization, *the earth's rotation*, which appears in the previous sentence, premodified by its agent (*the earth's*). In other cases, the agent may not be present in the same phrase or sentence but it can be easily retrieved from the context applying our shared knowledge of the world. This is the case of (73)

> (73) By **making Tryal** it will appear, that the Square of Saturn's Periodical Time, is to the Square of Jupiter's, as the Cube of Saturn's Diſtance from the Sun, is to the Cube of Jupiter's Diſtance (Gordon, 1726, p. 113; emphasis added).

where the agent of *making tryal* is meant to be interpreted as the spectator of the planets and stars. It is obvious that the agent must be an animated entity with the ability of trying something. Similarly, this animate agent has to be able to making tryals about celestial bodies, which excludes animals and unskilled people and narrows our focus down to skilled spectators of the heavens. Again, there is no reason to believe that stylistic nominalizations have particular mystification properties as the agent of most

nominalizations is either present in the same sentences or are ostensively inferential from shared knowledge of the world.

Concerning those stylistic nominalizations with an overt agent, the general trend is to accommodate it in the premodifying field (17%) and only 6% of them –the lowest percentage of agency inclusion across typologies– include an *of-construction*. Those premodified agents have mostly the shape of a possessive determiner. In (143)

> (143) The ſecond Satellite performs **his Revolution** in 3 Days, 13 Hours, 13 Minutes, 41 Seconds, 54 Thirds, 26 Fourths (Hodgson, 1749, p. 92; emphasis added).

the possessive *his* offers a reduplication of the agent *Satellite* that had already been made explicit in the grammatical subject of the sentence. This construction is exclusive of stylistic nominalizations as it plays with the structure of light-verb constructions.

Up to this point, overt exclusion of agents in the phrase governed by a nominalization does not seem to imply the existence of an agentless construction. It has been shown that agents are most times embedded in other parts of the sentence or the paragraph or they can be inferred without great complication. Term nominalizations, however, seem to complicate the panorama. They are in principle agentless constructions: 83% do not have an agent in their phrase and in most cases, their agent cannot be retrieved from the text or inferred from our shared knowledge of the world. It is true, however, that the situation is not as frustrating as it seems: term nominalizations are not as different from other typologies as it might seem. There are also multiples instances where the agent can be retrieved from the context. In (144)

> (144) FIGURE AND DIMENSIONS OF THE EARTH. A fluid mass rotating about an axis, and of which the particles attract one another with intensities varying inversely as the square of their distances apart, will assume the form of an oblate spheroid. **Its** axis of **rotation** will be both the shortest and a principal axis of figure (Bartlett, 1855, p. 21; emphasis added).

the agent of the nominalization *rotation* is present although it does not modify it. The possessive *its* is, formally, the premodifier of the nominalization's head. In spite of the apparent formal complexity, our cognitive knowledge of the world simplifies things out: The Earth rotates around an axis. This is common knowledge and it is therefore uneconomic to make it explicit every time we refer to it.

It is true, however, that 5,244 nominalizations, that is the 62% of the total number of all nominalizations found in this study, have no explicit agent. Such a big number must not be ignored. Of all these agentless nominalizations, 67% are term nominalizations. This is derived from a powerful feature of nominalizations as focalizers of information. Focalizers provide information to the reader to guide the interpretation proccess. they focus discourse on a particular topic by eliminating distractions, which in the case of nominalizations are not only agents but all verbal valencies. This sentence topic is in accordance with the mental model that the writer has made about the process and they try to replicate it onto the decoding process. That way, besides guiding the interpretation of the message, they help the reader avoid processing costs of disambiguating possible alternative meanings. In the case of conditioned nominalizations, agent omission may be more difficult because of grammar needs. In term nominalizations the situation is almost inverse because these nominalizations are,

among all typologies, the ones that share less properties with verbs and are closer to other nouns that do not express a process.

The frequencies of circumstance inclusion in nominalization NPs –shown in figure 60– are, as might be expected, pretty low. They range from 17% of circumstance inclusion in conditioned nominalizations to 30% of stylistic nominalizations.

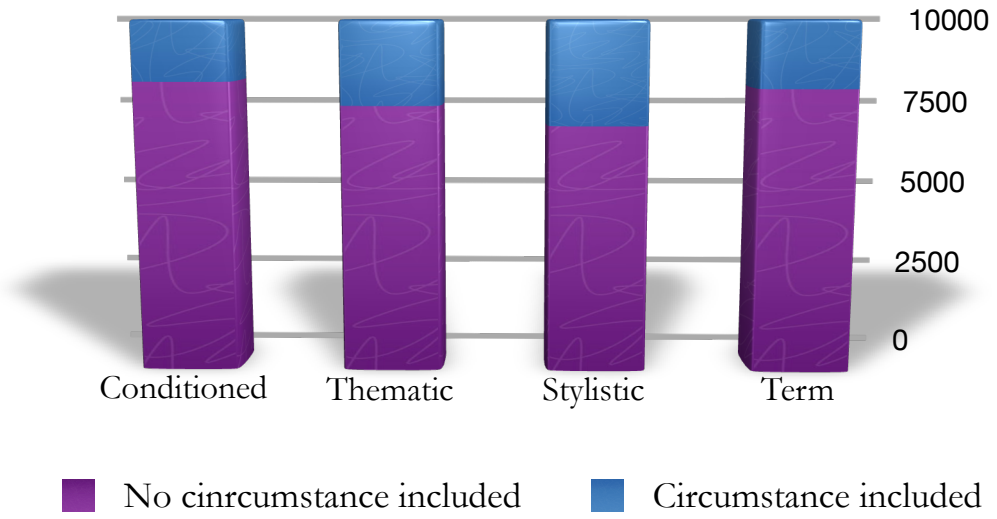


Figure 60: *Distribution of circumstance inclusion according to typology (NF 10,000).*

Circumstances are normally more peripheral elements than agents. In the case of verbal realizations, circumstances are normally expressed through optional adjuncts, whereas agents are normally obligatory elements. According to a transformalist view, its inclusion in the nominalization NP is even more unlikely to happen. Given its optional character in the VG, it would be more expectable to transform other verbal valencies first. Nominalizations, however, are different ways of encoding information and thanks to their structural flexibility, they allow the highlighting of the circumstances surrounding the process without the inclusion of other semantic elements.

Stylistic nominalizations are the best example to prove to what degree nominalizations and not mere verbal transformations, but independent ways of communicating processes. The high percentage of circumstance inclusion in this typology (30%) is due to the high frequency of *motion* in the corpus. To find a synonym for such a common word, authors may restore to a light-verb construction with *to have* or *to perform*. As analyzed in previous sections, verb choice depends on circumstance inclusion. Hence, in (130)

> (130) [...] he [the Sun] ſeems to **have an annual motion in the heavens**, and to riſe and ſet continually in different parts of them (Bonnycastle, 1786, p. 29; emphasis added).

the inclusion of a temporal and spatial circumstance in which the process took place, has selected *to have* for the construction. This claim may seem erratic but once the agent is included in the NP, the verb is changed[109]:

> (138) The Ecliptic is that great Circle in whoſe Plane the Earth **performs its annual Motion** round the Sun (Fuller, 1732, p. 6; emphasis added).

This selection rule, which applies consistently in all stylistic nominalizations with circumstance specification, shows to what extent nominalizations are different ways of expressing a process that give preference to focusing strategies concerning information transmission over grammar constraints.

---

[109] There are instances of stylistic nominalizations featuring to have, a possessive determiner and an adjective not expressing a circumstance. In (145)

> (145) [...] that is, **having their proper motion** from Weſt to Eaſt, or according to the order of the Signs [...] (Costard, 1767, p. 285; emphasis added).

proper does not express a circumstance about the process, but rather an attribute of motion, which has acquired nominal semantic features. In all stylistic nominalizations analyzed, adjectives expressing a circumstance in the process expressed in the nominalization are in charge of verb selection.

The low percentage of circumstance inclusion in both conditioned (17%) and term (19%) nominalizations is caused by two different factors. In the case of conditioned nominalizations, formal restrictions concerning grammatical context may be the main reason for low circumstance inclusion. This may be also due to the result of a focusing process that highlights the meaning of the process while setting back other neighboring elements, as, in this case, circumstances about the process. The same focusing process can be used to explain the low circumstance inclusion in term nominalizations (19%). In this case, given the semantic distance between verbs and term nominalizations, it is expectable to find less specifications about circumstances surrounding the process. Term nominalizations are nouns expressing process meanings in a reified way. Additionally, by reducing circumstances to a minimum, they are focusing on the process itself. For this reason, they can be considered functional guidelines for information processing. They provide instructions for information processing and limit the potential ambiguity of the utterances.

To sum up briefly, in this section, the application of extralinguistic variables did not provide important insights into typology variation. The retaining of semantic verbal properties and the distance of the nominalization from verbal realizations seem to motivate more formal and functional differences across typologies. After all, in this study, all typologies are conceived as belonging to a continuum that links verbal and nominal encodings. As a result, the rates of inclusion of agents and circumstances is higher in thematic nominalizations because they are semantically closer to the way in which verbs express processes. That is reflected in a higher rate of pre- and postmodification. In the case of term nominalizations, which are completely reified nouns, the tendency is just the opposite. The form of stylistic nominalizations is also

quite peculiar. Normally being part of a light-verb construction fulfilling stylistic functions, these nominalizations tend to avoid Germanic roots and suffixes and favor the inclusion of APs indicating circumstances in which the process took place. Syntactic function and the existence of a semantically-emptied verb are indeed powerful indicators of typological variation. Hence, whereas stylistic nominalizations normally function as direct objects, subject positions are usually fulfilled by thematic nominalizations. With their wordy modifiers, these nominalizations usually constitute a rewriting of a previous sentence and allow the advancement of discourse by grounding previous information and giving way to the introduction of new premises in the second part of the sentence. In the case of term nominalizations, adjuncts and modifiers are exceedingly the preferred syntactic function. In this case, phrases are shorter and most of the agents and circumstances about the process are not included. They have been reified and usually act as focalizers that turn the attention into the process itself and introduce it as a circumstance in another process. All in all, different structural differences across typologies seem to be the consequence of different functional needs.

5. Conclusions

The preceding chapters have dealt thoroughly with the causes and consequences of the Scientific Revolution and how it affected language change and the development of the scientific register in English in the eighteenth and nineteenth centuries (chapter one). Considering that nominalizations are a well-known scientific discourse marker in English, I have also reviewed the main theories about them and covered their chief structural and functional features (chapter two). Subsequently I have commented on the work tools and the corpus of texts used for analysis (chapter three). Using them, I have extracted all the deverbal nominalizations formed by suffixation from the corpus, disambiguated and classified them according to a series of linguistic and extralinguistic variables, which were consequently analyzed (chapter four). As a result of the method of study and the analysis of variables I have been able to reach a series of conclusions that will be presented below.

First, it has been shown that the claim made by Halliday (2004, p. 172) that the use of nominalizations and other scientific discourse markers is a result of an ongoing

process that started 400 or 500 years ago is consistent with data found in *CETA*. The tendency observed in frequency rates along the two centuries shows a slight, steady increase. The analysis carried out shows an increase of 58% in the mean frequency of nominalization use per sample text. Besides, not only the number of tokens, but also the number of types augmented. After discarding an explanation based on the complexity in the topic of texts, the cause of this increment could be the consolidation of nominalizations as markers of scientific register in English. Once established as scientific markers, the use of nominalizations became associated with the new model of science and the practices performed by the discourse community of the new men of science. The institutionalization movement that took place in the eighteenth century and the professionalization of science that happened in the nineteenth century strengthened the position of science in society and, consequently, also that of its practitioners and its language.

Second, texts produced by female writers tend to include a slightly superior number of nominalizations. Given the mainstream conception of the period that considered women unable to achieve abstract thought, the abundant use of nominalizations, which have been noted as abstraction facilitators (Downing, 1997; Eggins, 1994), by women makes it evident that they were indeed capable of abstract intellectual abilities. Additionally, as a result of their active role as readers of science their writing may have been influenced consistently including all the stylistic guidelines established by the discourse community. Thus, in spite of being banned from official science, their frequent use of nominalizations points out that they knew and mastered the standard register. Although I am aware of the fact that the number of texts written by women in my corpus is scarce, data results contradicted my expectations. In some way

it could be argued that these data demonstrate to what degree our idea of the exclusion of women from science is also part of a rhetorical device rooted in the ideology of the initiators of the Scientific Revolution. I am not implying that female exclusion from science did not exist but rather that this situation was originated and exacerbated as a result of the misogyny of the Revolution's initiators. Bacon and his followers established a standard for the scientific method that included, among other features, plainness of style. As we saw in section 3.1.2.2, the allusion to this feature is common in prefaces but in most cases this reference to plain style is a mere rhetorical device: it shows that the author knows and agrees with the main tenets of the new methodology, even if then he does not always obey them completely. The very adoption of nominalizations as scientific discourse markers is a proof of that, as they increase the level of ambiguity. Similarly, my claim is that the harsh situation for women to access official science was based on Bacon's hostility towards women. He considered science a masculine, active endeavor and believed it should have a new language that would differ from the gallant, poetic style of women (Schiebinger 1995). His attitude was shared by the early members of the Royal Society and misogyny pervaded official science in anglophone countries until the twentieth century. It is not true, however, that women abandoned science. After the seventeenth century, some women continued to be devoted to scientific activities as translators, assistants, practitioners and hosts of scientific salons. A minority even published their works, which, according to data analysis, were at the same level than their male counterparts, at least as far as the communication of science is concerned.

Third, there seems to be no correlation between nominalization use and the geographical variable. Even if figure 31 shows that the percentage of nominalizations

found in American texts is higher, this is a reflection of an unequal evolution of astronomy in both continents. Thus, whereas in England astronomy was a well-rooted discipline that served as a motor for the Scientific Revolution in the seventeenth century, science in colonial America depended mainly on England. However, after 1776, astronomy in the U.S. soon grew up to attain the development level of European science. After all, American colonist scientists were usually educated in Europe and the newly born country only needed some institutional support, which was granted only after it became independent. This historical situation is illustrated in *CETA*, as no text written by an American author from the eighteenth century has been included. Thus, the high frequency found in American texts is rather a reflection of the establishment of nominalization as a scientific discourse marker that was going on at the time. The study of nominalization use in each continent also shows that the place of education of authors did not have a direct effect on the language. This finding reinforces the power of the scientific discourse community as a compact entity with clear guild codifications. Despite geographical distance, members of the scientific community regarded themselves as a closed community and by means of language, they were able to distinguish members of the community from applicants and general society.

One further conclusion is that formality seems to be correlated with nominalization use. From all the nominalizations found in the texts, 63% of them were belong to formal texts and letters whereas research articles were the text-types with a higher number of nominalizations. In this study I have put all the text-types provided by *CETA* compilers in two groups considering the intended addressee of texts. My expectation was that texts aimed at a specialized audience would include a higher frequency of nominalization use because: a) nominalizations increase the level of

abstraction and ambiguity of texts due to valency reduction processes that hinder the presence of agents and their particular, more flexible way of organizing information about the process; and b) they are scientific discourse markers and their use indicates belongingness to the discourse community. My hypothesis was confirmed by data although the margin between both groups is not as high as I had expected and this may be due to the fact that nominalizations are not exclusively a discourse marker; they are also basic features in adult language and therefore they are present in all types of texts.

Fourth, the Romance component is more productive to form nominalizations. After etymologic study of both roots and suffixes, it was found that the Germanic component was less frequent and it also tended to decrease slightly over the period studied. Of the seven suffixes included for study, *-ion* had 65% of the total number of occurrences and its variation rate across the two centuries was 3.35%. On the other hand, the second most productive suffix, *-ing* represented only 15% and, most importantly, it tended to decrease with a mean rate of 7.37%. Root etymology follows the same trend, as 94% of occurrences had a Romance origin and, as we saw in figure 39, their frequency augmented notably over the two centuries. Concerning combinations of roots and suffixes, considering figure 40 it became clear that same-origin combinations are preferred. After the strong 94% of purely Romance nominalizations, Germanic pairings are the most repeated with 5% of the total number of occurrences and blends of roots and suffixes with different origins are minimal in number. The Romance component is indubitably the dominant element and this may be a consequence of linguistic practices of the scientific discourse community. One of the main changes that brought the Scientific Revolution was the introduction of the vernacular as a replacement for Latin as the language of scientific communication. The

main aim of the new scientists was to transform the understanding of the universe and by talking and writing in their own language they could achieve that faster. However, all scientists were still proficient in written Latin because that was the language in which they had received their education, so it is normal that they resorted to this language for the creation of new words. In some way, the high frequency of Romance nominalizations shows the Revolution was not as disruptive as it may have seemed and that it involved a natural evolution that encompassed old practices.

Fifth, the structural analysis of the elements in nominalization NPs showed that there are no restrictions concerning the type of modifiers that nominalizations can take. In the premodifying field, determiners, and more specifically, definite articles were the most repeated element. However, possessive use was especially indicative and it was shown that the person choice was a valuable indicator of the function of the nominalization NP. Thus, first-person possessives are normally dictated by stylistic concerns, as the author decides to claim his/her visibility in the text. Second-person possessives, on the other hand, are a result of text-type and they are mostly found in dialogues and textbooks, where the author wants to reduce the distance with the reader by addressing directly to him/her. Finally, third-person possessives, the most frequent type, have textual implications as they condense information about agency and refer to previous parts of the text in which the process and its agent were expressed through a finite sentence. Concerning postmodification, PPs, as we saw in figure 47, constitute 84% of the total, mainly because among the possibilities for postmodification, this is the most efficient way of encoding information about the process. PPs can contain much information, but, looking at data in figure 49, it became clear that agency was the most common element. This finding is directly linked to the next conclusion.

Nominalizations have a flexible way of communicating information about agents and circumstances. Here, two contrasting realities collide. Even if it is true that, as we saw in figures 50 and 51, the 38% of agency and the 21% of circumstance inclusion are relatively low numbers, their importance is key in the understanding of nominalizations as focalizers of information because by means of its particular way of organizing information about processes in the phrase, they may highlight or hinder information about agents, objects and circumstances and consequently reinforce or diminish the degree of attention on the process. Concerning agency inclusion, data analysis has shown that with regard to verbal realizations, nominalizations are more flexible and admit agents not only as postmodifying PPs, but also as possessive determiners and even APs. In many cases, the agent is not contained in the nominalization NP because it can be retrieved from previous parts of the text or from our shared knowledge of the world. In these cases, agency exclusion is not a mystificatory tool, but rather an economical device complying with Grice's Maxim of Quantity. The 21% of circumstance inclusion, on the other hand, can be considered fairly surprising. Reckoning that for a great deal of scholars nominalizations are transformations of verbal form or meaning, circumstances, which are normally encoded as optional modifiers in the VP, would be the first element likely to be excluded in NPs. Thus, the percentage of circumstance inclusion in nominalization NPs in *CETA* is indicative that nominalizations indeed include and organize information in their own way to suit the structural and functional needs of the text.

Seventh, syntax delimits the functional implications of nominalizations in texts. In this field my expectation of a majority of nominalizations fulfilling a subject position was reversed, as we saw in figure 43 by the finding that 55% of occurrences functioned

as adjuncts. This is perhaps the most surprising outcome of data analysis, as the traditional description of nominalizations in scientific register focused on the subject function, as it was believed (Halliday, 2004) that in this position nominalizations serve as encapsulators of previous discourse that facilitate the advancement of discourse and the assimilation of complex ideas. subject position is indeed popular and 20% of nominalizations were found fulfilling this function. However, the methodology used in this study enabled a fresh look at the syntactic variable by highlighting the importance of nominalizations functioning in peripheral sentence positions. The functional implications of these nominalizations expand the concept of textual coherence. They are powerful deictical devices that draw temporal, spatial and causal connections between processes. Unlike verbs, they can be highly economical, as all of their modifiers are optional. Apart from building on cohesion and coherence, the conciseness of completely reified nominalizations can have very positive implications for the processing and assimilation of information, as all attention is drawn to the process and other participants and circumstances are left in the background.

Eighth, nominalizations are not idle features and they are usually complemented by semantically-emptied verbs. At this point, the claim made by Halliday (2004: 155) that verbs in English scientific register were enduring a grammaticalization process that turned them into connectors between processes presented in nominal groups seems grounded. It has been found that many nominalizations in subject and subject complement positions were accompanied by semantically-emptied verbs. However, more interestingly, it was also detected that in many cases those semantically-emptied verbs were combined with nominalizations in Object positions, forming light-verb constructions. The textual function of these constructions was associated with stylistic

concerns. Even if they may be also understood as a special type of synonymy that avoids lexical repetition, the use of these constructions increases the degree of complexity and abstraction and, consequently, defines the intended audience. Again, we find that a great deal of nominalization use is dictated not only by textual concerns but also by linguistic practices of a particular discourse community.

Ninth, there are significant differences according to nominalization typology and the prevalence of term nominalizations highlights the value of nominalizations as points of reference within texts. For this study, I have created a typology of lexical nominalizations that assessed not only the proximity to nouns and verbs but also the functionality in texts. Thus, thematic nominalizations are close to verbs, they tend to include lengthy modifiers that encode information about the process and they serve as encapsulators of information that organize and add coherence to texts. Given the significant degree of scholarly attention that they have received, my expectation was to find a high number of them but, surprisingly they represented 30% of the total number of occurrences and, as we saw in figure 52 they were outnumbered by term nominalizations. Stylistic nominalizations are considered a type of guild codification because they add structural complexity and their main function is to meet stylistic concerns. They constitute 8% of the total number of occurrences and are the less frequent typology. Finally, conditioned and term nominalizations are the result of a process of reification that can be seen not only in their tendency to include very few, short modifiers but also in their progressive loss of the semantics of the process. Conditioned nominalizations, on the one hand, are normally dictated by structural needs of the sentence in which they are inserted. Term nominalizations, on the other, function as indexes of information, establishing points of reference in the text and facilitating

information processing and knowledge assimilation. They represent 50% of the total number of nominalizations found in *CETA*. This finding reinforces the cognitive value of the role of nominalizations in texts.

Extralinguistic and morphological variables did not provide significant differences among typologies. Concerning chronology, except conditioned nominalizations, all typologies had a tendency to increase. The exception of conditioned nominalizations is justified by the fact that their use is normally dictated by structural needs of the text and consequently there will always be a constant rate of them. The variables of place of education and sex of author produced practically identical results and in the case of the text-type variable the only noticeable difference is the slightly superior tendency that stylistic nominalizations show to be found in formal texts. This trend, however, was extremely expectable because formal texts are, in theory, more prone to include stylistic concerns that would complicate the structure of the text as a way of delimiting the degree of proficiency in the topic and its writing standards that intended addressees are expected to possess.

Finally, agent inclusion, syntactic function and the existence of a semantically-emptied verb are the most useful criteria for typology discrimination. There are appreciable differences in the way different typologies accomodate information about their agents, as we saw in figure 59. As it might be expected, thematic nominalizations, due to their proximity to verbal encodings, show the highest degree of agent inclusion (70%), whereas the completely reified term nominalizations barely incorporate it in the phrase. If agency inclusion was determined by the closeness regarding verbs and nouns, the syntactic variable defines the functionality of different typologies in texts. In this light, most conditioned and term nominalizations were found in adjunct positions.

Again, the functionality of reified nominalizations functioning as indexes of information that enforce text coherence and facilitate the assimilation of information is revealed. On the other end of the scale, stylistic nominalizations were mainly found as objects of semantically-emptied verbs in light-verb constructions. Finally, the semantics of verbs was also important as, according to data analysis, there is a high tendency that these verbs take a thematic nominalization as their subject –or rather the opposite.

My hope is that this doctoral dissertation may have clarified the role of nominalizations in English scientific register in the eighteenth and nineteenth centuries. This is indeed a very interesting topic that can have a direct impact on our understanding of both nominalizations and the way we communicate science. Future work on this topic may involve a reassessment of data analysis to include clausal nominalization and other types of lexical nominalization not formed by suffixation. Similarly, contrastive studies with other *CC* subcorpora could be of general utility, because they would broaden the scope of our comprehension of this topic. Nevertheless I am hopeful that the conclusions I have reached after data analysis may inspire future work on the role of nominalizations in scientific register.

Bibliography

Primary Sources

Adams, George (1777). *A Treatise describing the construction and explaining the use of celestial and terrestrial globes. Designed to illustrate, in the most easy and natural manner the phoenomena of the Earth and Heavens, and to shew that correspondence of the two spheres. With a great variety of astronomical and geographical problems. London: printed and sold by the author.*

Bartlett, William Holmes Chambers (1855). *Elements of natural philosophy. IV Spherical astronomy*. New York: A. S. Barnes & Burr.

Bonnycastle, John (1786). *An introduction to astronomy in a series of letters*. London: J. Johnson.

Bradford, Duncan (1845). *The wonders of the heavens, being a popular view of astronomy, including a full illustration of the mechanism of the heavens, embracing the Sun, Moon, and Stars, etc.* Boston: Otis, Broaders and Company.

Brewster, David (1811). *Ferguson's astronomy explained upon Sir Isaac's principles: with notes and supplementary chapters*. Edinburgh: J. Ballantine & Co.

Bryan, Margaret (1797). *A compendious system of astronomy in a course of familiar lectures; in which the principles of that science are clearly elucidated, so as to be intelligible for those who have not studied the Mathematics. Also trigonometrical and celestial problems, with a key to the ephemeris, and a vocabulary of the terms of science used in the lectures, which latter are explained agreeably to their application in them. London: M. Bryan.*

Charlton, Jasper (1735). *The Ladies Astronomy and Chronology in four parts*. London: T. Gardner.

Chauvenet, William (1871). *A manual of spherical and practical astronomy, embracing the general problems of spherical astronomy, the special applications to nautica astronomy, and the theory and use of fixed and portable astronomical instruments.* Philadelphia: Lippincott.

Clerke, Agnes M. (1893). *A popular history of astronomy during the nineteenth century*. London: Adam & Charles Black.

Costard, George (1767). *The history of astronomy, with its application to geography, history, and chronology; occasionally exemplified by the globes*. London: J. Lister.

Croll, James (1889). *Stellar evolution and its relation to geological time*. New York: Appleton & Co.

Curson, Henry (1702). *The theory of sciences illustrated, or the grounds and principles of the seven arts: Grammar, logick, rhetorick, musick, arithmetick, geometry, astronomy.* London: R. Smith.

Darwin, George Howard (1880). On the secular changes in the elements of the orbit of a satellite revolving about a tidally distorted planet. *Philosophical Transactions of the Royal Society of London for the year MDCCCLXXX. 171*, Part II, 864-873.

Ewing, John (1809). *A plain elementary and practical system of natural experimental philosophy; including; astronomy and chronology.* Philadelphia: Hopkins & Earle.

Ferguson, James (1756). *Astronomy explained upon Sir Isaac Newton's principles and made easie to those who have not studied mathematics*. London: printed for, and sold by the author.

Fuller, Samuel (1732). *Practical astronomy, in the description and use of both globes, orrery and telescopes.* Dublin: S. Fuller.

Garland, Landon C. (1838). Address on the utility of astronomy. *Southern Literary messenger; devoted to every department of literature and the fine arts*, *4*(2), 123-130.

Gordon, Gordon (1726). *An introduction to geography, astronomy, and dialling. Containing the most useful elements of the said sciences, adapted to the meanest capacity, by the description and uses of the terrestrial and celestial globes. With an introduction to chronology.* London: J. Senex.

Gummere, John (1822). *An elementary treatise on astronomy in two parts. The first, containing a clear and compendious view of the theory. The second, a number of practical problems.* Philadelphia: Kimber and Sharpless.

Harris, John (1719). *Astronomical dialogues between a gentleman and a lady*. London: T. Wood.

Herschel, John F. W. (1833). *A treatise on astronomy.* London: A. Spottiwood.

Hill, John (1754). *Urania: Or, a compleat view of the heavens.containing the antient and modern astronomy, in form of a dictionary: Illustrated with a great number of figures.*London: T. Gardner.

Hodgson, James (1749). *The theory of Jupiter's satellites*, *with the* Construction and use of the tables for computing their eclipses. London: W. Mount & T. Page.

Lacy, John (1779). *The universal system: Or mechanical cause of all the appearances and movements of the visible heavens:shewing the true powers which move the earth and planets in their central and annual rotations. With a dissertation on comets, the nature, cause, matter, and use of their tails, and the reasons of their long trajectories: likewise an attempt to prove what it is that moves the sun round its axis.* London: J. Buckland.

Long, Roger (1742). *Astronomy, in five books* (I). Cambridge: R. Long.

Loomis, Elias (1868). *A treatise on astronomy.* New York: Harper.

Lowell, Percival (1895). Mars: III Canals. *The Atlantic Monthly*, *76*, 106-119.

Luby, Thomas (1828). *An introduction treatise to physical astronomy.* London: Baldwin and Cradock.

Mitchel, Ormsby Macknight (1860). *Popular astronomy. A concise elementary treatise on the sun, planets, satellites and comets.* New York: Phinney, Blekeman & Mason.

Morden, Robert (1702). *An introduction to astronomy, geography, navigation, and other mathematical sciences made easie by the description and uses of the coelestial and terrestrial Globes*. London: R. Morden & R. Smith.

Nicholson, William (1782). *An introduction to natural philosophy* (I). London: J. Johnson.

Olmstead, Denison (1841). *Letters on astronomy, addressed to a lady*. Boston: Marsh, Capen, Lyon, and Webb.

Philips, William (1817). *Eight familiar lectures on astronomy. An introduction to science for the use of young persons and others not conversant with the mathematics.* New York: J. Eastburn & Co.

Small, Robert (1804). *An account of the astronomical discoveries of Kepler: Including an historical review of the sistems which had successively prevailed before his time*. London: J. Mawman.

Steele, Joel Dorman (1874). *Fourteen weeks in descritive astronomy.* New York: A. S. Barnes.

Stewart, Matthew (1761). *Tracts, physical and mathematical: Containing, an explication of several important points in physical astronomy and a new method for ascertaining the sun's distance from the earth*. London: A. Millar & J. Nourse.

Vince, Samuel (1790). *A treatise on practical astronomy*. Cambridge: J. Merrill, J. Nicholson & W. Lunn.

Watts, Isaac (1726). *The knowledge of the heavens and the earth made easy: Or, the first principles of astronomy and geography explain'd by the use of globes and maps*: *With a solution of the common problems by a plain* scale and Compasses as well as by the globe. London: J. Clark, R. Hett & R. Ford.

Whewell, William (1858). *The plurality of worlds. With an introduction by Edward Hitchcock.* New York: Sheldon, Blakeman & co.

Whiston, William (1715). *Astronomical lectures* (I). London: R. Senex & W. Taylor.

Wilson, Alexander (1774). Observations on the solar spots. *Philosophical transactions (1683-1775)*, *XLIV*, 1-19.

Young, Charles A. (1880). Recent progress in solar astronomy. *The Princeton review.* January-June, 88-104.

Works Cited

Albentosa Hernández, José Ignacio (1997). La sustantivación en el discurso científico en lengua inglesa. *CAUCE. Revista de Filología y su Didáctica, 20-21*, 329-344.

Albentosa Hernández, José Ignacio & Moya Guijarro, Arsenio Jesús (2000). La reducción del grado de transitividad de la oración en el discurso científico en lengua inglesa. *Revista Española de Lingüística, 30*(1), 445-468.

Alexiadou, Artemis (2001). *The functional structure in nominals: Nominalizations and ergativity*. Amsterdam/Philadelphia: John Benjamins.

Algarotti, Francesco (1739). *Il Newtonianismo per le dame, ovvero dialoghi sopra la luce e i colori*. Milano: A spese di Giambatista Pasquali libraro e stampatore di Venezia.

Alonso Ramos, Margarita (1994-1995). Hacia una definición del concepto de colocación: De J. R. Firth a l. A. Mel'čuk. *Revista de Lexicografía, 1*, 9-28.

Alonso Ramos, Margarita (2004). *Las construcciones con verbos de apoyo*. Madrid: Visor.

Applebaum, Wilbur (Ed.). (2000). *Encyclopedia of the Scientific Revolution: From Copernicus to Newton*. New York: Garland Publishing.

Astell, Mary (1697). *A serious proposal to the ladies, for the advancement of their true and greatest interest. In two parts. By a lover of her sex.* London: Printed for Richard Wilkin at the King's Head in St. Paul's Church-Yard.

Atkinson, Dwight (1998). *Scientific Discourse in sociohistorical context: The Philosophical Transactions of the Royal Society of London 1675-1975*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bacon, Francis (1985 [1696]). *The essayes or cousels, civil and moral.* Oxford: Clarendon Press.

Bailey, Richard (1996). *Nineteenth-century English*. Ann Harbor: The University of Michigan Press.

Baker, Cristopher (Ed.). (2002). *Absolutism and the Scientific Revolution, 1600-1720: A biographical dictionary.* Westport, CT & London: Greenwood Press.

Baker, Paul (2006). *Using corpora in discourse analysis*. London/ New York: Continuum.

Bally, Charles (1951). *Traité de stylistyque française*. Paris: Klincksieck.

Banks, David (2001). Analyse des discours spécialisés: Le domaine anglais. *Revue Française de Linguistique Appliquée, VI*(2), 7-16.

Banks, David (2003). The evolution of grammatical metaphor in scientific writing. In Anne Marie Simon-Vandenbergen, Miriam Taverniers & Louise Ravelli (Eds.), *Grammatical Metaphor: Views from Functional Linguistics* (pp. 127-147). Amsterdam: John Benjamins.

Banks, David (2005a). Emerging scientific discourse in the late seventeenth century. *Functions of Language, 12*(1), 65-86.

Banks, David (2005b). On the historical origins of nominalized process in scientific text. *English for Specific Purposes, 24*(3), 347-357.

Banks, David (2007). L'évolution de la phrase en anglais scientifique. In David Banks (Ed.), *La coordination et la subordination dans le texte de spécialité* (pp. 203-221). Paris: L'Harmattan.

Banks, David (2008). *The development of scientific writing: Linguistic features and historical context.* London: Equinox.

Barber, Charles (1976). *Early Modem English*. Edinburgh: Edinburgh University Press.

Barber, Charles (1993). *The English language: A historical introduction*. Cambridge: Cambridge University Press.

Baroni, Marco & Bernardini, Silvia (2004). BootCaT: Bootstrapping corpora and terms from the web. In Marco Baroni & Silvia Bernardini (Eds.), *Proceedings of LREC 2004* (pp. 1313-1316). Bologna: GEDIT.

Bartholomew, Michael & Morris, Peter (1991). Science in the Scottish Enlightenment. In David Goodman & Colin A. Russell (Eds.), *The rise of scientific Europe: 1500-1800* (pp. 279-304). Sevenoaks, UK: Hodder & Stoughton Limited.

Beal, Joan C. (2012). Late modern English in its historical context. In Isabel Moskowich, & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The writing of science between 1700 and 1900* (pp. 1-14). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bello, Iria (2010a). Nominalizations in astronomical texts in the eighteenth century. In Isabel Moskowich-Spiegel, Begoña Crespo, Inés Lareo & Paula Lojo (Eds.), *Language windowing though corpora* (pp. 75-89). A Coruña: Universidade da Coruña.

Bello, Iria (2010b). A diachronic study of nominalizations in astronomy and philosophy texts. Paper delivered at the *6th International Contrastive Linguistics Conference*, Freie Universität Berlin, 30/09-02/10/2010. (Unpublished manuscript).

Bello, Iria & Müller, Catharina (2011). Españoles holgazanes vs. alemanes aplicados: Metáfora cognitiva y configuración conceptual de la expresión de "trabajo" en español y alemán. Paper delivered at *FRASESPAL 2011, Internationale Tagung zur Kontrastiven Phraseologie Deutsch-Spanisch/Galicisch*, Universidad de Santiago de Compostela, 24-25/11/2011. (Unpublished manuscript).

Bensaude-Vincent, Bernadette & Blondel, Christine (Eds.). (2008). *Science and spectacle in the European Enlightenment.* Burlington, VT: Ashgate Publishing Company.

Bentley, Michael (Ed.). (1997). *Companion to historiography*. London & New York: Routledge.

Betham-Edwards, Matilda (1880). *Six life-studies of famous women*. London: Griffith & Farran.

Bhat, Shankara D. N. (1994). *The adjectival category.* Amsterdam: John Benjamins.

Biber, Douglas (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas (1993). Representativeness in corpus design. *Literary and Linguistics Computing, 8*(4), 243-257.

Biber, Douglas, Finegan, Edward, & Atkinson, Dwight (1994). ARCHER and its challenges: Compiling and exploring a representative corpus of historical English registers. In Udo Fries, Peter Schneider & Gunnel Tottie (Eds.), *Creating and using English language corpora. Papers from the 14th International Conference*

on English language research on computerized corpora, Zurich 1993 (pp. 1-13). Amsterdam: Rodopi.

Billig, Michael (2008). The language of critical discourse analysis: The case of nominalization. *Discourse Society, 19*, 783-800.

BootCaT. Simple utilities to bootstrap corpora and terms from the web. Retrieved October 19, 2012, from http://bootcat.sslmit.unibo.it/?section=hom20

Blakemore, Diane (1987). *Semantic constraints on relevance.* Oxford: Blackwell.

Briones, Susana, Fortuny, Liliana, Sastre, Silvia & Botto de Porcovi, Martha (2003). Grammatical metaphors in scientific English. *The ESPecialist, 24*(2), 131-142.

Brush, Stephen G. (1988). *The history of modern science: A guide to the second scientific revolution, 1800-1950*. Ames, Iowa: Iowa State University Press.

Burke, Peter (2000). *A social history of knowledge: From Gutenberg to Diderot*. Cambridge, Malden, MA: Polity Press.

Burke, Peter (2004). *What is cultural history?* Cambridge: Polity Press.

Burton-Roberts, Noel (1998). *Analysing sentences. An introduction to English syntax* (2nd ed.). London & New York: Longman.

Butterfield, Herbert (1965). *The origins of modern science*. New York: The Free Press.

Cadematori, Yanina, Parodi, Giovanna & Venegas, René (2006). El discurso escrito y especializado: Caracterización y funciones de las nominalizaciones en los manuales técnicos. *Literatura y Lingüística, 17*, 243-265.

Calvin, William H. (1996). *Como piensan los cerebros*. Madrid: Editorial Debate.

Camiña Rioboo, Gonzalo (2010). New nouns for new ideas. In María LLuisa Gea-Valor, Isabel García, & María José Esteve (Eds.), *Linguistic and Translation Studies in scientific Comunications* (Vol. 86) (pp.157-176). Bern/Berlin: Peter Lang.

Camiña Rioboo, Gonzalo (2012). Accounting for observations of the heavens in the 18th century. new nouns to explain old phenomena. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The writing of science between 1700 and 1900* (pp. 93-121). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Camiña Rioboo, Gonzalo (2013). *Word formation in the scientific register of early Modern English: A corpus-based approach*. (Doctoral dissertation). Available from Repositorio da Universidade da Coruña.

Camiña Rioboo, Gonzalo, Esteve, María José, & Lareo, Inés (2012). A study of nouns and their provenance in the astronomy section of the *Coruña Corpus of Scientific English Writing*: Some preliminary considerations". In José Luis Cifuentes Honrubia, Adelina Gómez Gonzalez-Jover, Antonio Lillo Buades, José Mateo Martínez & Francisco Yus Ramos (Eds.), *Los caminos de la lengua. Estudios en homenaje a Enrique Alcaraz Varó* (pp. 537-547). Alicante: Universidad de Alicante.

Castiglione, Baldassare (1528). *Il libro del cortegiano*. In Venetia: Nelle case d'Aldo Romano & d'Andrea d'Asola.

Cavendish, Margaret (1664). *Philosophical letters*. London.

Chomsky, Noam (1957). *Syntactic structures*. London: Mouton.

Chomsky, Noam (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.

Chomsky, Noam (1970). Remarks on nominalization. In A. Jacobs Roderick & Peter S. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 184-221). Waltham, MA: Ginn and Co.

Ciapuscio, Guiomar (1992). Impersonalidad y desagentivación en la divulgación científica. *Lingüísitca española actual, XIV*(2), 183-205.

Coffin, Caroline (2006). *Historical discourse: The language of time, cause and evaluation*. London: Continuum.

Coley, Noel (1991). Science in seventeenth-century England. In David Goodman & Colin A. Russell, *The rise of scientific Europe: 1500-1800* (pp. 197-226). Sevenoaks, UK: Hodder & Stoughton Limited.

Comrie, Bernard & Thompson, Sandra (1985). Lexical nominalizations. In Timothy Shopen (Ed.), *Language typology and syntactic description* (vol. 3) (pp. 349–398). Cambridge: Cambridge University Press.

Crespo, Begoña (2004a). General survey of the growth of scientific culture. A historical approach. In Elizabeth Woodward et al. (Eds.), *About culture* (pp. 57-65). A Coruña: Universidade da Coruña.

Crespo, Begoña (2004b).The scientific register in the history of English: A corpus-based study. *Studia Neophilologica, 76*, 125-139.

Crespo, Begoña (2011). Persuasion markers and ideology in eighteenth century philosophy texts (CEPhiT). *Revista de Lenguas para Fines Específicos, 14/15*, 199-228.

Crespo, Begoña (2012a). Astronomy as scientific knowledge in modern England. In Isabel Moskowich, & Begoña. Crespo (Eds.), *Astronomy 'playne and simple': The writing of science between 1700 and 1900* (pp. 15-34). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Crespo, Begoña (2012b). *Change in life, change in language: A semantic approach to the history of English*. Bern: Peter Lang.

Crespo, Begoña (in press 2014a). Involved in writing science: Nineteenth-century women in the Coruña Corpus. *International Journal of Corpus Linguistics*.

Crespo, Begoña (in press 2014b). The use of persuasion strategies by women writing science. *Australian Journal of Feminist Studies*.

Crespo, Begoña (in press 2014c). The Woman is said to be the weaker Vessel: Scientists of both sexes in the late Modern period. *Journal of Historical Pragmatics.*

Crespo, Begoña & Moskowich, Isabel (2005). Latin forms in vernacular scientific writing: Code-switching or borrowing?. In Rod MacConchie et al. (Eds.), *Selected proceedings of the 2005 symposium on new approaches in English historical lexis (Hel-Lex)* (pp. 51-59). Somerville, MA: Cascadilla Press.

Crespo, Begoña & Moskowich, Isabel (2009). The limits of my language are the limits of my world: The scientific lexicon from 1350 to 1640. *SKASE Journal of Theoretical Linguistics, 6*(1), 45-58.

Crespo, Begoña, Puente, Luis, Bello, Iria & Lojo, Paula (2012). A corpus of women scientists in the 18th and 19th centuries: Their lives and works. Poster presented at *CILC 12*. (Unpublished manuscript).

Crombie, Alistair (1974). *Historia de la ciencia: De San Agustín a Galileo* (II). *siglos XIII-XVII* (2nd ed.). Madrid: Alianza Universidad.

Crystal, David (1992). *An encyclopedic dictionary of language and languages.* Oxford: Oxford University Press.

Curzan, Anne (2010). Language and gender: How different are men and women. Talk delivered at the University of A Coruña on October 13, 2010. (Unpublished manuscript).

Díaz Hormigo, María Tadea (2011). Sobre los denominados sustantivos deverbales de acción. *Documentos de trabajo. Gramática comparada y descriptiva 20*/6. doi:e-archivo.uc3m.es/handle/10016/11534

van Dijk, Teun A. (1988). *News as Discourse*. Hillsdale: Erlbaum.

van Dijk, Teun A. (2006). Ideology and discourse analysis. *Journal of Political Ideologies, 11*(2), 115-140.

van Dijk, Teun A. (2008). Critical discourse analysis and nominalization: Problem or pseudo-problem? *Discourse Society, 19*, 821-828.

Dik, Simon C. (1985). Formal and semantic adjustment of derived constructions. In A. Machelt Bolkenstein et al. (eds.), *Predicated and Terms in Functional Grammar* (pp. 1-28). Dordrecht: Foris Publications.

Dik, Simon C. (1997). *The theory of functional Grammar (II): Complex and derived constructions*. Berlin: Mouton de Gruyter.

Downing, Angela (1992). *A university course in English grammar*. New York: Prentice Hall.

Downing, Angela (1997). Encapsulating Discourse Topics. *Estudios Ingleses de la Universidad Complutense, 5*, 147-168.

Downing, Angela (2000). Nominalization and topic management in leads and headlines. In Eija Ventola (Ed.), *Discourse and community: Doing functional linguistics* (pp. 355-378). Tübingen: Gunter Narr.

Eberle, Kurt, Faasz, Gertrud & Ulrich, Heid (2011). Approximating the disambiguation of some German nominalizations by use of weak structural, lexical and corpus information. *Procesamiento del Lenguaje Natural, 46*, 67–75.

von Eckardt, Barbara (1993). *What is cognitive science?* Cambridge, MA: The MIT Press.

Eger, Elizabeth (1999). *Bluestocking feminism: Writings of the Bluestocking circle, 1738-1785*. London: Pickering & Chatto.

Eggins, Suzanne (1994). *An introduction to systemic functional linguistics*. London: Pinter.

Evans, Vyvyan, Bergen, Berger & Zinken, Joerg (2007), The cognitive linguistics enterprise: An overview. In Evans, Vyvyan, Bergen, Berger & Zinken, Joerg (Eds.), *The cognitive linguistics reader* (pp. 2-36). London: Equinox Publishing Ltd.

Fairclough, Norman (1992). *Discourse and social change*. Cambridge: Blackwell publishers.

Fauconnier, Gilles (1994). *Mental spaces: Aspects of meaning construction in natural language.* New York: Cambridge University Press.

Fauconnier, Gilles & Turner, Mark (2002). *The way we think. Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Feingold, Mordechai (1989). The universities and the Scientific Revolution: The case of England. In Robert P.W. Visser. (Eds.), *New trends in the history of science: Proceedings of a conference held at the University of Utrecht* (pp. 29-48). Amsterdam: Rodopi.

Fillmore, Charles (1975). *Frame Semantics*. Berkeley: University of California.

Firth, John R. (1957). Modes of meaning. *Papers in Linguistics 1934-1951*, 190-215. London: Oxford University Press.

Fitzmaurice, James (2011). Cavendish, Margaret, duchess of Newcastle upon Tyne (1623?–1673). In *Oxford dictionary of national biography*. Retrieved May 5, 2011 from http://www.oxforddnb.com/view/article/4940

Fowler, Roger (1991). *Language in the news: discourse and ideology in the press.* London & New York: Routledge.

Freedberg, David (2002). *The eye of the lynx: Galileo, his friends, and the beginnings of modern natural history.* Chicago & London: The University of Chicago Press.

Gallais-Hamonno, Janine (1981). The development of English as the language of science. *Le Langage Et l'Homme, 47*, 17-23.

Ghani, Rayid, Jones, Rosie, & Mladeni'c, Dunja (2003). Building minority language corpora by learning to generate web search queries. *Knowledge and Information Systems, 7*(1), 56-83.

Girju, Roxane, Nakov, Preslav, Nastase, Vivi, Szpakowicz, Stan, Turney, Peter & Yuret, Deniz (2009). Classification of semantic relations between nominals. *Language Resources and Evaluation, 43*(2), 105–121.

Givón, Talmy (1995). *Functionalism and grammar*. Amsterdam: John Benjamins.

Gómez López, Susana (2013). De la mímesis a la representación: Empirismo y lenguaje en los orígenes de la ciencia moderna. *Revista de Filosofía, 38*(1), 53-77.

Goodman, David & Russell, Colin A (Eds.), *The rise of scientific Europe: 1500-1800*. Sevenoaks, UK: Hodder & Stoughton Limited.

Gopnik, Myrna (1972). *Linguistic structure in scientific texts*. The Hague: Mouton.

Görlach, Manfred (1991). *Introduction to early Modern English.* Cambridge: Cambridge University Press.

Görlach, Manfred (1994). *The linguistic history of Engish*. London: Macmillan.

Görlach, Manfred (1997). *The linguistic history of English. An introduction.* Hampshire: MacMillan Press.

Görlach, Manfred (1999). *English in nineteenth-century England: An introduction*. Cambridge: Cambridge University Press.

Görlach, Manfred (2001). *Eighteenth-century English.* Heidelberg: Winter.

Gotti, Maurizio. (1992). The development of a scientific language in the 17th century. In Carmen Nocera Avila, Nicola Pantaleo & D. Pezzini (Eds.), *Early modern English: Trends, forms and texts. Papers read at the IV National Conference of history of English, Catania 2-3 may 1991* (pp. 319-343). Fasano: Schena editore.

Gotti, Maurizio (2001). The experimental essay in early Modern English. *European Journal of English Studies, 5*(2), 221-239.

Gotti, Maurizio (2011). The development of specialized discourse in the *Philosophical Transactions*. In Irma Taivitsainen & Päivi Pahta (Eds.), *Medical writing in early Modern English* (pp. 204-220). Cambridge: Cambridge University Press.

Gotti, Maurizio (2012). Variation in academic texts. In Maurizio Gotti (Ed.), *Academic identity traits: A corpus-based investigation* (pp. 23-42). Bern: Peter Lang.

Gotti, Maurizio (2013). The formation of the Royal Society as a community of practice and discourse. In Joanna Kopaczyk & Andreas Jucker (Eds.), *Communities of practice in the history of English* (pp. 269-285). Amsterdam: John Benjamins.

Graesser, Arthur C., Mcnamara, Danielle S. & Louwerse Max M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text. In Anne P. Sweet & Catherine E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York: Guildford Publications.

Greene, John C. (1954). Some aspects of American astronomy 1750-1815. *Isis, 45*(4), 339-358.

Greene, John C. (1984). *American science in the age of Jefferson.* Ames, Iowa: The Iowa State University Press.

Grefenstette, Gregory & Teufel, Simone (1995). Corpus-based method for automatic identification of support verbs for nominalizations. In Steven Abney & Erhard Hinrichs (Eds.), *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics, Dublin, Ireland* (pp. 98-103). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Grice, Paul (1975). Logic and conversation. In Peter Cole & Jerry Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

Grimshaw, Jane (1990). *Argument structure*. Cambridge: MIT Press.

Gross, Alan (1990). *The rhetoric of science*. Cambridge, MA: Harvard University Press.

Guillén Calve, Ignacio (1998). The textual interplay of grammatical metaphor on the nominalizations occurring in written medical English. *Journal of Pragmatics, 30*, 363-385.

Gutiérrez Rodilla, Bertha María (2003). La historia del lenguaje científico como parte de la historia de la ciencia. *Asclepio, 55*(2), 7-26.

Hall, Alfred Rupert (1954). *The Scientific Revolution 1500-1800: The formation of the modern scientific attitude*. London/NY/Toronto: Longmans.

Halliday, Michael A.K. (1966). Lexis as linguistic level. In Charles Bazell, John Firth, John Catford & Michael A. K. Halliday (Eds.), *In memory of J.R. Firth* (pp. 148-162). London: Longman.

Halliday, Michael A. K. (1985). *An introduction to functional grammar.* London: Continuum.

Halliday, Michael A. K. (2004). On the language of physical science. In Jonathan Webster (Ed.), *The language of physical science* (pp. 162-178). London: Continuum.

Halliday, Michael A. K. & Hassan, Ruqaiya (1976). *Cohesion in English*. Harlow: Longman.

Halliday, Michael A. K. & Martin, James (1993). *Writing Science*. London: The Falmer Press.

Halliday, Michael A. K. & Matthiessen, Christian M. (2004). *An introduction to functional grammar* (3rd ed.). London: Arnold.

Harcstark-Myers, Sylvia (1990). *The Bluestocking circle: Women, friendship and life of the mind in eighteenth-century England*. Oxford: Clarendon Press.

Hazout, Ilan (1995). Action nominalizations and the lexicalist hypothesis. *Natural Language and Linguistic Theory, 13*(3), 355-404.

Henry, John (2002). *The Scientific Revolution and the Origins of Modern Science.* New York: Palgrave.

Heyvaert, Liesbet (2003). *A cognitive-functional approach to nominalization in English*. Berlin: Mouton de Gruyter.

Heyvaert, Liesbet (2008). On the constructional semantics of gerundive nominalizations. *Folia Linguistica, 42*(1-2), 39-82.

Heyvaert, Liesbet (2010). A cognitive-functional perpective on deverbal nominalization in English: Descriptive findings and theoretical ramifications. In Monika Rathert

& Artemis Alexadiou (Eds.), *Semantics of Nominalizations across Languages and Frameworks* (pp. 51-82). Berlin: Mouton de Gruyter.

Hickey, Raymond (2003). *Corpus presenter software for language analysis with a manual and a corpus of Irish English as sample data*. Amsterdam: John Benjamins.

Hill, Christopher (1965). *Intellectual origins of the English Revolution*. Oxford: Oxford University Press.

Hopper, Paul & Thompson, Sandra (1980). Transitivity in grammar and discourse. *Language, 56*(2), 251-294.

Hull, Richard D. & Gómez, Fernando (2000). Semantic interpretation of deverbal nominalizations. *Natural Language Engineering, 6*(2), 139–161.

Inhelder, Bärbel & Piaget, Jean (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Inkster, Ian (1982). Advocates and audience: Aspects of popular astronomy in England, 1750-1850. *Journal of the British Astronomical Association, 92*(3), 119-123.

Jackendoff, Ray (1975). Morphological and semantic regularities in the lexicon. *Language, 51*(3), 639-671.

Jakobson, Roman (1981). *Selected writings. III, Poetry of grammar and grammar of poetry.* The Hague: Mouton.

Kilgariff, Adam & Grefenstette, Gregory (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics - Special Issue on Web as Corpus, 29*(3), 333-347.

Kilgariff, Adam, Rychlý, Pavel, Smrz, Pavel, & Tugwell, David (2004). The Sketch Engine. *Proceedings of EURALEX, Lorient, France* (pp. 105-116).

Koptjevskaja-Tamm, Maria (1993). *Nominalizations*. London: Routledge.

Koptjevskaja-Tamm, Maria (2003). Action nominal constructions in the languages of Europe. In Frans Plank (Ed.), *Noun Phrase Structure in the Languages of Europe* (pp. 723–761). Berlin & New York: Mouton de Gruyter.

Krämer-Friedrich, Sybille (1988). The 'Universal Thinking Machine', or on the genesis of schematized reasoning in the seventeenth century. In Imre Hronszky, Márta Fehér, Balázs Dajka (Eds.), *Scientific knowledge socialized: Selected proceedings of the 5th Joint International Conference on the History and Philosophy of Science* (pp. 179-191). Dordrecht: Reidel.

Kytö, Merja (2012). Manual to the diachronic part of the Helsinki Corpus of English Texts: Coding conventions and lists of source texts. Retrieved November 20, 2012, from http://icame.uib.no/hc/

Labov, William (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Lakoff, George & Johnson, Mark (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, George (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lakoff, George (2007). The contemporary theory of metaphor. In Vyvyan Evans, Berger Bergen & Joerg Zinken (Eds.), *The cognitive linguistics reader* (pp. 264-315). London: Equinox Publishing Ltd.

Langacker, Ronald (1987). *Foundations of cognitive grammar, Volume I, Theoretical prerequisites*. Stanford, California: Stanford University Press.

Lapata, Mirella (2002). The disambiguation of nominalizations. *Computational Linguistics, 28*(3),357-388.

Lareo, Inés (2006). Colocaciones verbo-nombre en inglés moderno tardío. Algunos aspectos de su comportamiento y evolución. (Doctoral dissertation). Universidade da Coruña.

Lareo, Inés (2008). Analysing a type of collocation. Make-complex predicates in 19th century science and fiction. *Revista Canaria de Estudios Ingleses, 57*, 165-180.

Lareo, Inés (2009a). Make-collocations in Nineteenth-Century Scientific English. *Studiá Neophilologica, 81*(1), 1-16.

Lareo, Inés (2009b). El *Coruña Corpus*. Proceso de compilación y utilidades del *Corpus of English Texts on Astronomy (CETA)*. Resultados preliminares sobre el uso de los predicados complejos en *CETA*. In Pascual Cantos Gómez & Aquilino Sánchez Pérez (Eds.), *A survey on corpus-based research* (pp. 267-280). Murcia: Asociación Española de Lingüística de Corpus. .

Lareo, Inés (2011). Using complex predicates in the writings on astronomy in nineteenth-century English. Exploitation of the corpus of english texts on astronomy. *Revista de Lenguas para Fines Especificos, 17*, 229-252.

Lareo, Inés (2012). A corpus-driven analysis of complex predicates in the 18th century scientific writings in *CETA*. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The writing of science between 1700 and 1900* (pp. 155-180). Amsterdam/Philadelphia: John Benjamins

Lareo, Inés & Esteve, María José (2008). 18th century scientific writing: A study of 'make' complex predicates in the *Coruña Corpus*. *ICAME Journal, 32*, 69-96.

Lareo, Inés & Montoya Reyes, Ana (2007). Scientific writing: Following Robert Boyle's principles in experimental essays-1704 and 1998. *Revista Alicantina De Estudios Ingleses, 20*, 119-137.

Lees, Robert B. (1960). *The grammar of English nominalizations*. The Hague: Mouton.

Lehmann, Christian (1988). Towards a typology of clause linkage. In John Haiman & Sandra Thompson (Eds.), *Clause Combining in Grammar and Discourse* (pp. 181–225). Amsterdam: John Benjamins.

Lehman, Christian (2002). *Thoughts on grammaticalization*. Erfurt: Seminar für Sprachswissenschaft der Universität.

Lemke, Jay L. (1995). *Textual Politics*. London: Taylor & Francis.

Levin, Carole et al. (2000). *Extraordinary women of the medieval and Renaissance world. A biographical dictionary*. Westport, Connecticut: Greenwood Press.

Lorenzano, César (2011). El lenguaje de la ciencia. *Arbor, 187*(747), 15-24 doi: 10.3989/arbor.2011.747n1002

Lureau, Maxence (2014). Le coaching strategico-linguistique: Vers une science du changement? (Doctoral dissertation). Université de Paris Ouest Nanterre-La Défense/Università di Roma-La Sapienza. (Unpublished manuscript)

Lyons, John (1977). *Semantics, Vol. 1*. Cambridge: Cambridge University Press.

Mackenzie, J. Lachlan (1985). Nominalization and valency reduction. In A. Machtelt Bolkenstein, Casper de Groot & J. Lahlan Mackenzie (Eds.), *Predicates and terms in functional grammar* (pp. 28-47). Dordrecht: Foris.

Mackenzie, J. Lachlan (1986). Aspects of nominalization in English and Dutch. *Working Papers in Functional Grammar, 15*, 1-32.

Mackenzie, J. Lachlan (1987). Nominalization and basic constituent ordering. In Johan van der Auwera & Louis Goossens (Eds.), *Ins and Outs of Predication* (pp. 93–106). Dordrecht: Foris.

Mackenzie, J. Lachlan (1996) English nominalizations in the layered model of the sentence. In B. Devriendt, L. Goossens & J. Van der Auwera (Eds.) *Complex Structures: A functionalist perspective* (pp. 325-355). Berlin & New York: Mouton de Gruyter.

Mackenzie, J. Lachland (2007). Double-possessive nominalizations in English. In C. S. Butler, R. Hidalgo Downing and J. Lavid (eds.) *Functional perspectives on grammar and discourse: Papers in honour of Angela Downing* (pp. 217-232). Amsterdam: Benjamins.

Maier, Robert (2007). Discourse and cultural transformation. In Shi-xu (Ed.), *Discourse as cultural struggle* (pp. 17-28). Hong Kong: Hong Kong University Press.

Malchukov, Andrej (2004). *Nominalization/Verbalization: Constraining a typology of transcategorial operations*. Munich: Lincom.

Malchukov, Andrej (2006). Constraining nominalization: Function-form competition. *Linguistics, 44-45*, 973-1008.

Marinkovich Ravena, Juana (2005). La nominalización como un recurso de cohesión léxica en los manuales de la formación técnico-profesional. *Revista de LIngüística Teórica y Aplicada, 43*(1), 23-44.

Martin, James R. (1993). Literacy in science: Learning to handle text as technology. In Michael A.K. Halliday & James R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 166-202). Pittsburgh: University of Pittsburgh Press.

Martin, James R. (2008). Incongruent and proud: De-vilifying "nominalization". *Discourse and Society, 19*(6), 801-810.

Martin, James R. & Veel, Robert (Eds.). (1998). *Reading science: Critical and functional perspectives on discourses of science*. London: Routledge.

Maynard, Senko K. (1994). The centrality of thematic relations in Japanese text. *Functions of Language,1*(2), 229-260.

McEnery, Tony, & Wilson, Andrew (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Mel'čuk, Igor, Arbatchewsky-Jumaire, Nadia, Elnitsky, Leo, & Iordanskaja, Lidija (1984). *Dictionaire explicatif et combinatoire du français contemporain*. Montréal: Presses de l'Université de Montréal.

Mel'čuk, Igor & Zholkovsky, Alexander (1970). Towards a functioning mening-text model of language. *Linguistics, 57*, 10-47.

Millward, Celia M. (1988). *A biography of the English language*. New York: Harcourt Brace.

Monge, Félix (1970). Los nombres de acción en español. In A. Rosetti (Ed.), *Actele celui de al XII-lea Congres Internacional de Linguistică si Filologie Romanică, 15.4-20.4 1968* (I), (pp. 961-972). Bucharest: Editura academiei republicii socialiste România.

Monge, Félix (1977). -Ción, -sión, -zón y -ón: Función y forma en los sufijos. In *Estudios ofrecidos a Emilio Alarcos Llorach (con motivo de sus 25 anos de docencia en la Universidad de Oviedo)* (II), (pp. 155-165). Oviedo: Servicio de publicaciones de la universidad de Oviedo.

Moreno Cabrera, Juan Carlos (1987). *Fundamentos de sintaxis general*. Madrid: Síntesis.

Moskowich, Isabel (2001). Morfología flexiva del inglés moderno. In Isabel de la Cruz & J. Martín (Eds.), *Lingüística histórica inglesa* (pp. 624-654). Barcelona: Ariel.

Moskowich, Isabel (2010a). To lerne sciences touching nombres and proporciouns: The proportion of affixation in early scientific writing. *English Studies, 39*(1), 39-52.

Moskowich, Isabel (2010b). Morphologically complex nouns in English scientific texts after Empiricism. *Linguistik Online, 43*(3).

Moskowich, Isabel (2012). *CETA* as a tool for the study of modern astronomy in English. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The writing of science between 1700 and 1900* (pp. 35-57). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Moskowich, Isabel (2013). Eighteenth-century female authors: Women and science in the *Coruña Corpus of English Scientific Writing*. *Australian Journal of Linguisitics, 33*(4) 467-487.

Moskowich-Spiegel, Isabel & Crespo, Begoña (2007). Presenting the *Coruña Corpus: A collection of samples for the historical study of English scientific writing*. In Javier Pérez-Guerra, Jorge Bueno Alonso, Dolores González-Álvarez & Esperanza Rama Martínez (Eds.), *Of varying language and opposing creed: New insights into late Modern English* (pp. 341-357). Berlin/New York: Peter Lang.

Moskowich, Isabel & Crespo, Begoña (2012). *Astronomy 'playne and simple': The writing of science between 1700 and 1900*. Amsterdam: John Benjamins.

Moskowich Isabel, Lareo Inés, Camiña Rioboo Gonzalo & Crespo, Begoña. (Compilers). (2012). *A corpus of English texts on astronomy (CETA)* (CD ROM). Amsterdam/Philadelphia: John Benjamins.

Moskowich, Isabel & Monaco, Leida Maria (2014). Abstraction as a means of expressing reality: Women writing science in late Modern English. In Maurizio Gotti & D. S. Giannoni (Eds.), *Corpus analysis for descriptive and pedagogical purposes* (pp. 203-224). Bern: Peter Lang.

Moskowich, Isabel & Parapar, Javier (2008). Writing science, compiling science. The *Coruña Corpus of English Scientific Writing*. In María Jesús Lorenzo Modia (Ed.), *Proceedings from the 31st AEDEAN Conference* (pp. 531-544). A Coruña: Universidade da Coruña.

Murdin, Lesley (1985). *Under Newton's shadow: Astronomical practices in the seventeenth century*. Bristol & Boston: Adam Hilger Ltd.

Muste. (2010). Introduction to the Research Group for Multidimensional Corpus-Based Studies in English (MuStE). Retrieved November 15, 2012, from http:// www.udc.es/grupos/muste/

Nevalainen, Terttu, Raumolin-Brunberg, Helena, Jukka, Keränen, Nevala, Minna, Nurmi, Arja & Palander-Collin, Minna (Compilers). (1998). *Corpus of early English correspondence*. Helsinki: University of Helsinki.

Newmeyer, Frederick J. (1971). The source of derived nominals in English. *Language, 47*(4), 786-796.

Newton, Isaac (1687). *Philosophiæ naturalis principia mathematica*. Londini: Jussu Societatis Regiæ ac typis Josephi Streater. Prostat apud plures bibliopolas.

Oostdijk, Nelleke (1991). *Corpus linguistics and the automatic analysis of English*. Amsterdam/Atlanta: Rodopi.

Oxford English Dictionary Online (1989). (2nd edition). Oxford: Clarendon Press.

Padó, Sebastian, Pennacchiotti, Marco & Sporleder, Caroline (2008). Semantic role assignment for event nominalisations by leveraging verbal data. In D. Scott & Hans Uszokoreit (Eds.), *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1, COLING '08* (pp. 665–672). Stroudsburg, PA: Association for Computational Linguistics.

Parapar, Javier & Moskowich, Isabel (2007). The Coruña Corpus Tool. *Procesamiento De Lenguaje Natural, 38*, 289-290.

Parker, David (2011). Algarotti, Francesco, Count Algarotti in the Prussian nobility (1712–1764). In *Oxford dictionary of national biography*. Retrieved May 6, 2011 from http://www.oxforddnb.com/view/article/68123

Pérez-Guerra Javier, Bueno Alonso Jorge, González-Álvarez Dolores & Rama Martínez, Esperanza (Eds.). (2007). *Of varying language and opposing creed: New insights into late modern English*. Berlin/New York: Peter Lang.

Peris, Aina (2012). *Nominalizaciones deverbales: Denotación y estructura argumental.* (Doctoral dissertation). Available from Centre de Llenguatge i Computació.

Perry, Ruth (2011). Astell, Mary (1666–1731). In *Oxford dictionary of national biography*. Retrieved May 5, 2011 from http://www.oxforddnb.com/view/article/814.

Picallo, Carme (1999). La estructura del sintagma nominal: Las nominalizaciones y otros sustantivos con complementos argumentales. In Ignacio Bosque & Violeta

Demonte (Eds.), *Gramática descriptiva de la lengua española* (I) (pp. 363–393). Madrid: Espasa Calpe.

Poulain de la Barre, François (1679). *De l'egalité des deux sexes, discours phisique et moral, où l'on voit l'importance de se défaire des Préjuguez*. (2nd ed.). Paris: chez Antoine Dezallier.

Prince, Alan & Smolensky, Paul (2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

Pumfrey, Stephen (1997). "The Scientific Revolution". In Bentley, Michael (Ed.), *Companion to Historiography* (pp. 293-306). London & New York: Routledge.

Pustejovsky, James (1995). *The Generative Lexicon*. Cambridge, Massachussets: The Mit Press.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan (1985 (1993)). *A Comprehensive grammar of the English language.* New York: Longman.

Ravelli, Louise J. (1988). Grammatical metaphor: An initial analysis. In Erich Steiner & Robert Veltman (Eds.), *Pragmatics, discourse and text. Some systemically-inspired approaches* (pp. 133-147). London: Pinter.

Rissanen, Matti, Kytö, Merja, Kahlas-Tarkka, Leena, Nevanlinna, Saara, Taatvitsainen, Irma, Nevalainen, Terttu, et al. (Compilers), (1991). *The Helsinki Corpus of English Texts*. Helsinki: University of Helsinki.

Robb, T. (2003). Google as a quick 'n' dirty corpus tool. *Teaching English as a Second or Foreign Language, 7*(2).

Roberts, Gerrylynn (1991). Scientific academies across Europe. In David Goodman & Colin A. Russell (Eds.), *The rise of scientific Europe: 1500-1800* (pp. 227-252). Sevenoaks, UK: Hodder & Stoughton Limited.

Rossiter, Margaret W. (1982). *Women scientists in America: Struggles and strategies to 1940.* Baltimore: John Hopkins University Press.

Rothenberg, Marc (1981). Organization and control: Professionals and amateurs in American astronomy, 1899-1918. *Social Studies of Science, 11*(3) 305-325.

Rothenberg, Marc (1985). History of astronomy. *Osiris, 1,* 117-131.

Russell, Colin A. (1991). The reception of Newtonianism in Europe. In David Goodman & Colin A. Russell (Eds.), *The rise of Scientific Europe: 1500-1800* (pp. 253-278). Sevenoaks, UK: Hodder & Stoughton Limited.

Saggi di naturali esperienze fatte nell'Accademia del cimento sotto la protezione del serenissimo principe Leopoldo di Toscana e descritte dal segretario di essa accademia. (1667) Florence: Giuseppe Cocchini.

Schiebinger, Londa (1995 [Original French 1989]). *The mind has no sex?: Women in the origins of modern science.* Cambridge, MA: Blackwell Publishers.

Schmied, Josf (1994). The Lampeter corpus of early Modern English tracts. In Merja Kytö, Matti Rissanen & S. Wright (Eds.), *Corpora across the centuries* (pp. 81-89). Amsterdam/Atlanta: Rodopi.

Sharoff, Serge (2006). Creating general-purpose corpora using automated search engine queries. In Marco Baroni & Silvia Bernardini (Eds.), *Wacky! working papers on the web as corpus* (pp. 63-98). Bologna: GEDIT.

Siemund, Rainer & Claridge, Claudia (1997). The Lampeter corpus of early Modern English. *ICAME, 21,* 61-70.

Siemund, Rainer (2012). The Lampeter Corpus of early Modern English Tracts. Retrieved November 20, 2012, from http://khnt.hit.uib.no/icame/manuals/LAMPETER/LAMPHOME.HTM

Siloni, Tal (1997). *Noun phrases and nominalizations. The syntax of DPs*. Dordrecht/ Boston/ London: Kluwer Academic Publishers.

SketchEngine. (2012). *SketchEngine*. Retrieved October 19, 2012, from http:// bootcat.sslmit.unibo.it/?section=hom

Snyder, Laura J. (2009). Hypotheses in 19th century British philosophy of science: Herschel, Whewell, Mill. In Michael Heidelberger & Gregor Schiemann (Eds.), *The significance of the hypothetical in the natural sciences* (pp. 59-76). Berlin: Mouton De Gruyter.

Sperber, Dan & Wilson, Deirdre (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford/Cambridge: Blackwell Publishers.

Sperber, Dan & Wilson, Deirdre (2006). Relevance theory. In G. Ward and L. Horn (Eds.), *Handbook of pragmatics* (pp. 606-632). Oxford: Blackwell.

Sprat, Thomas (1667). *The history of the Royal-Society of London, for the improving of natural knowledge*. London: Printed by T.R. For J. Martyn at the Bell.

Steans, Raymond (1970). *Science in the British colonies of America*. Urbana, IL: University of Illinois Press.

Stewart, Aubrey & Long, George (Eds.). (1880). Plutarch's *Lives*. London: George Bell & Sons.

Stewart, Larry (1992). *The rise of public science*. Cambridge: Cambridge University Press.

Stewart, Larry (2008). The laboratory, the workshop, and the theatre of experiment. In Bernadette Bensaude-Vincent & Christine Blondel (Eds.), *Science and spectacle in the European Enlightenment* (pp. 11-24). Burlington, VT: Ashgate Publishing Company.

Sušinskienė, Solveiga (2004). Grammatical metaphor in scientific discourse. *Linguistics: Germanic and Romance studies, 54,* 76-83.

Sušinskienė, Solveiga (2009). Nominalization as a micro-structural item of English scientific discourse. *Respectus Philologicus, 16*(21), 84-92.

Sušinskienė, Solveiga (2010a). Textual functions of nominalizations in English scientific discourse. *Zmogus ir zodis 11*(3)*,* 58-63.

Sušinskienė, Solveiga (2010b). Nominalization as a cohesive device in British Newspaper Editorials. *Filologija / Philology: Scientific Publications, 15*, 142-150.

Sušinskienė, Solveiga (2012). Nominalization as a lexico-grammatical cohesive device in science popular texts. *Filologija / Philology: Scientific Publications*, *17*, 133-144.

Swales, John M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taatvitsainen, Irma (2001). Changing conventions of writing: The dynamics of genres, text types, and text traditions. *European Journal of English Studies, 5*(2), 139-150.

Taavitsainen, Irma (2005). On corpus linguistics: Computers and the history of English. In Isabel Moskowich-Spiegel & Begoña Crespo (Eds.), *Re-interpretations of English: Essays on language, linguistics and philology* (II), (pp. 325-345). A Coruña: Universidade da Coruña.

Taatvitsainen, Irma, & Pahta, Paivï. (1997). The corpus of Early English medical writing: Linguistic variation and prescriptive collocations in scholastic style. In Terttu Nevalainen, & Leena Kahlas-Tarkka (Eds.), *To explain the present: Studies*

in changing English language in honour of Matti Rissanen (pp. 209-225). Helsinki: Société Néophilologique.

Taavitsainen, Irma & Pahta, Päivi (Eds.). (2004). *Medical and scientific writing in late medieval English*. Cambridge: Cambridge University Press.

Taatvitsainen Irma., Pahta Paivï & Mäkinen Martti (Compilers). (2005). *Middle English Medical Texts.* (CD-ROM). Amsterdam: John Benjamins.

Taatvitsainen Irma, et al. (Compilers), (2010). *Early Modern English Medical Texts.* (CD-ROM). Amsterdam: John Benjamins.

TEI. (2012). *TEI: Text Encoding Initiative*. Retrieved October 15, 2012, from http:// www.tei-c.org/index.xml

Thagard, Paul (1996). *Mind: Introduction to cognitive science*. Cambridge, MA: The MIT Press.

The Ladies' diary, or complete Almanack, for the year of our Lord 1831, being the third after Bissextile, or Leap Year, containing new improvements in arts and sciences, and many Entertaining Particulars: Designed for the use and diversion of the Fair Sex. (1831). London: Printed for the company of stationers, by J. & C. Adlard, Bartholomew Close.

Tognini-Bonelli, Elena (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Trew, Tony (1979). What the papers say: Linguistic variation and ideological difference. In Roger Fowler, Bob Hodge, Gunther Kress & Tony Trew (Eds.), *Language and Control* (pp. 117-156). London: Routledge.

UNESCO. (1988). *Proposed international standard nomenclature for fields of science and technology*. Paris: UNESCO/ROU257 rev. 1.

University of Manchester. (2012). *ARCHER: A Representative Corpus of Historical English Registers*. Retrieved November 20, 2012, from http:// www.llc.manchester.ac.uk/research/projects/archer/

Valeika, Laimutis (1985). *An introduction to the linguistic analysis and synthesis of the text.* Vilnius: Vilnius University Press.

Varieng. (2012). Corpus of Early English correspondence. Retrieved November 20, 2012, from http://www.helsinki.fi/varieng/domains/CEEC.html

Varieng. (2012). Corpus of Early English medical writing. Retrieved November 20, 2012, from http://www.helsinki.fi/varieng/CoRD/corpora/CEEM/

Vázquez Orta, Ignacio (2006). A corpus-based approach to the distribution of nominalization in academic discourse. In María José Luzón, Silvia Murillo & Ana María Hornero (Eds.), *Corpus linguistics: Applications for the study of English* (pp. 399-416). Bern: Peter Lang.

Ventola, Eija (1996). Packing and unpacking of information in academic texts. In Eija Ventola & Anna Mauranen (Eds.), *Academic writing, intercultural and textual issues* (pp. 153-194). Amsterdam: John Benjamins.

Ventola, Eija (1999). Semiotic Spanning at conferences: Cohesion and coherence in and across conference papers and their discussions. In Wolfram Bublitz, Uta Lenk & Eija Ventola (Eds.), *Coherence in spoken and written discourse: How to create and how to describe it: Selected papers from the International Workshop on Coherence, Augsburg, 24-27 April 1997*. Amsterdam: John Benjamins.

Voss, Jürgen (1980). Die Akademien als Organisationsträger der Wissenchaften im 18. Jahrhundert. *Historische Zeitschrift, 23*(1), 43-74.

Webster, John (1654). *Academiarum examen, or the Examination of academies : Wherein is discussed and examined the matter, method and customes of academick and scholastik learning, and the insufficiency thereof discovered and laid open; as also some expedients proposed for the reforming of schools, and the perfecting and promoting of all kind of science / offered to the judgements of all those that love the proficiencie of arts and sciences, and the advancement of learning, by Jo. Webster*. London: Printed for Giles Calvert.

Wodak, Ruth & Meyer, Michael (Eds.). (2001). *Methods of critical discourse analysis*. London: Sage Publications.

Woolf, Virginia (1929). *A room of one's own*. London: published by Leonard & Virginia Woolf at the Hogarth Press.

Wu, Zongjie & Lu, Qinxia (2007). The discourse of Chinese medicine and westernization. In Shi-xu (Ed.), *Discourse as cultural struggle* (pp. 155-176). Hong Kong: Hong Kong University Press.

Yeomans, Donald K. (1977). The origin of North American astronomy: Seventeenth century. *Isis, 68*(3), 414-425.

Zizek, Slavoj (Ed.). (1994). *Mapping ideology*. London: Verso.

Zubizarreta, María Luisa (1987). *Levels of representation in the lexicon and in the syntax*. Foris. Dordrect.

Zucchi, Alessandro (1993). *The language of propositions and events*. Dordrecht/Boston/ London: Kluwer Academic Publishers.

Appendix A

Doctorado Internacional (2014)

Programa: Novas aproximacións aos estudos ingleses

Departamento de Filología Inglesa

Universidad de A Coruña

Directora: Begoña Crespo García

RESUMEN DE TESIS DOCTORAL

ON HOW *THE MOTION OF THE STARS* CHANGED THE LANGUAGE OF SCIENCE: A CORPUS-BASED STUDY OF DEVERBAL NOMINALIZATIONS IN ASTRONOMY TEXTS FROM 1700 TO 1900.

(DE COMO *EL MOVIMIENTO DE LAS ESTRELLAS* CAMBIÓ EL LENGUAJE DE LA CIENCIA: UN ESTUDIO BASADO EN CORPUS DE NOMINALIZACIONES DEVERBALES EN TEXTOS DE ASTRONOMÍA ESCRITOS ENTRE 1700 Y 1900).

Iria María Bello Viruega

La ciencia y el lenguaje de la ciencia son dos entidades indisolubles. La dificultad del lenguaje científico no se limita al nivel léxico, sino que también se aplica a una serie de estructuras gramaticales específicas (Halliday, 2004). Las nominalizaciones, uno de los marcadores del discurso científico más conocidos, son estructuras complejas que codifican procesos en sustantivos. Al nominalizar, una gran cantidad de información semántica se suele dejar fuera, lo que aumenta el grado de ambigüedad y la dificultad de descodificar correctamente la frase.

El objeto de la investigación de este estudio aborda el uso de las nominalizaciones deverbales formadas mediante sufijación en textos de registro científico escritos en inglés en los siglos XVIII y XIX. En este lapso de tiempo, varios cambios sociales afectaron la configuración de la ciencia en Europa y esto tuvo repercusiones en el lenguaje. Los empiristas establecieron una serie de pautas sobre la morfosintaxis, el léxico especializado y la estructura del texto para la presentación de los documentos de carácter observacional o experimental. Estas pautas estilísticas se convirtieron en un estándar de escritura después de la fundación de la Royal Society (Crespo, 2004, 2011). Las nominalizaciones se convirtieron en uno de estos marcadores distintivos y en consecuencia, su complejidad y frecuencia han aumentado progresivamente en inglés desde entonces.

El material corpus para este estudio fue tomado del *Corpus of English Texts on Astronomy (CETA)* (Moskowich et al., 2012). Este es uno de los subcorpus del *Coruña Corpus, A Collection of Samples for the Historical Study of English Scientific Writing*. *CETA* contiene dos textos por década y cada muestra contiene alrededor de 10.000 palabras, lo que hace un total de 800.000 palabras analizables. Otra de las herramientas de trabajo utilizadas ha sido la *Coruña Corpus Tool* (*CCT*), una herramienta de recuperación de la información diseñada específicamente para su uso conjunto con el corpus.

Entiendo la nominalización como una expresión lingüística de una representación conceptual de un proceso o estado de cosas en una forma nominal. Esta definición se basa en las premisas establecidas por Downing (1997, p. 147), quien considera que las situaciones y los procesos se pueden expresar lingüísticamente de dos maneras principales: cláusulas y nominalizaciones. Los lenguajes naturales tienen múltiples

recursos para expresar ideas similares. Esta flexibilidad, que puede acarrear importantes matices de significado, ha atraído la atención de lingüistas y hablantes, que ven reflejada la riqueza del lenguaje. Como resultado, los procesos se pueden expresar por medio de nominalizaciones, como en (1)

> (1) From whence it is gathered, that **the apparent progreſſive Motion of the Fixed Stars** hath gone forward one Degree towards the conſequent Signs, in about Seventy Years ſpace (Whiston, 1715, p. 14; énfasis añadido).

o a través de cláusulas con verbos conjugados, como en (2)

> (2) Aſtronomers know that not only the 12 Conſtellations of the Zodiac, but alſo **all the fix'd Stars move from the Weſt toward the Eaſt** about 50" in a Year, or one Degree in 71 Years, in Circles parallel to the Ecliptick (Watts, 1726, p. 34; énfasis añadido).

Obviamente, aunque en términos generales ambas opciones transmiten el mismo significado, cada una de estas dos codificaciones lingüísticas tiene una estructura diferente y cumple diferentes funciones en el texto. En (2) el verbo *move* controla la sintaxis de la frase entera a través de un sistema de valencias obligatorias y modificadores opcionales. Así, el agente (*Conſtellations of the Zodiac, but alſo all the fix'd Stars*) aparece como sujeto. La dirección del movimiento (*from the Weſt toward the Eaſt*) se hace explícita y también se da información sobre cómo (i*n Circles parallel to the Ecliptick*) y cuánto (*about 50″ in a Year, or one Degree in 71 Years*) las constelaciones y las estrellas se mueven. Del mismo modo, en (1) la nominalización *motion* también ejerce control sobre su frase, que a la vez se inserta en una oración más grande. La estructura no es tan rígida en este caso ya que, por definición, todos los elementos de la frase nominal con excepción del núcleo son opcionales. Esto permite

una disposición más compleja de los elementos. Así, en (1) la información sobre el agente (*of the Fixed Stars*), y cómo (*progressive*) el movimiento se lleva a cabo se encuentra dentro de los modificadores nominales. Esto sin embargo no implica la reducción de la cantidad de información ya que en la frase verbal se da también información sobre la dirección (*towards the conſequent Signs, in about Seventy Years ſpace*) y la duración (*one Degree)* del movimiento.

Funcionalmente, la expresión del proceso como un verbo en (2) se sitúa más cerca de la experiencia del hablante en términos de secuenciación cronológica y experiencia de la realidad ya que toda la información sobre el proceso se mantiene cerca del verbo. Sin embargo, la configuración a través de una nominalización permite presentar el proceso no como un simple relato de la realidad, sino como consecuencia del discurso anterior que viene a ser percibida como un objeto, dada su codificación como sustantivo.

La funcionalidad de las nominalizaciones ha sido el punto central en la investigación de muchos autores, sobre todo dentro de la escuela sistémico funcional (Banks, 2005a, 2005b; Guillén, 1998; Halliday, 1985, 2004; Ventola, 1996). Una breve reseña de las principales funciones realizadas por las nominalizaciones en los textos podría incluir: cohesión léxica (repetición y resumen); economía, concisión, condensación de la información y contextualización de la información (en relación con las estructuras de información), que se traduce en el avance del discurso. El dinamismo que las nominalizaciones confieren a la estructura temática de los textos a menudo se enfrenta a los efectos de la eliminación del agente, un tema ampliamente tratado por la escuela del análisis crítico del discurso (Billig, 2008; van Dijk, 2006, 2008; Fairclough, 1992; Fowler, 1991; Wodak y Meyer, 2001). Las nominalizaciones pueden silenciar

información sobre los agentes y las circunstancias, lo que puede dar lugar a ciertos efectos ideológicos. Sin embargo, desde un punto de vista cognitivo, las nominalizaciones responden a un modelo mental que el escritor ha hecho sobre el proceso que quiere comunicar y el contexto del proceso de comunicación. En este sentido, las nominalizaciones se convierten en pautas funcionales para el procesamiento de la información ya que proporcionan instrucciones y limitan la posible ambigüedad de las frases, ahorrando esfuerzos en el procesamiento (Blakemore, 1987). Para este estudio, he desarrollado una tipología de las nominalizaciones que tiene en cuenta características formales y funcionales:

1. **Nominalización condicionada**: por lo general viene dictada por las necesidades gramaticales del texto. Extremadamente concisa, suele ser la opción que escritores utilizan para condensar varios procesos en una sola frase y omitir elementos innecesarios. En (3)

> (3) The same comet, also, came very near the earth; so that, had its quantity of matter been equal to that of the earth, it would, by its **attraction**, have caused the earth to revolve in an orbit so much larger than at present, (...) (Olmstead 1841: 318; énfasis añadido).

el autor quiere presentar el proceso (atracción) como agente causante de otro proceso (que la Tierra gira en una órbita). Debido a que la información sobre el agente, objeto y circunstancias en las que el proceso de la atracción tiene lugar son fácilmente extraíbles del contexto y se derivan de nuestro conocimiento compartido del mundo, el escritor prefirió centrar la atención del lector en el proceso en sí mismo y presentarlo en forma de una nominalización .

2. **Nominalización temática**: cumple con todas las características descritas por los funcionalistas (Banks 2001, 2005a, 2005b): un proceso codificado en un grupo verbal (codificación congruente) y que funciona como rema se puede convertir en una nominalización (metáfora gramatical) en el tema de la siguiente frase. Así sirve como dispositivo de cohesión, repite y resume la información y constituye un punto de partida para que el avance del discurso. En (4)

> (4) but if the limbs be ſeperated, the object end is inclined to the quadrant, and muſt be **adjuſted** accordingly, and repeat the operation till the limbs coincide at both wires, and the **adjuſtment** is made (Vince, 1790, p. 15; énfasis añadido).

se consigue que el discurso avance convirtiendo el verbo *adjusted* en la nominalización *adjustment* al final de la oración. Semánticamente, este tipo de nominalización presenta una gran similitud con la codificación de los procesos por medio de cláusulas con verbo conjugado.

3. **Nominalización estilística**: la elección de la codificación nominal responde, en esta tipología, a criterios estilísticos. La complejidad estilística se considera una forma de "codificación gremial", un código que sólo los miembros de una comunidad dominan y que diferencia a los miembros experimentados de los novatos y del resto de la sociedad (Ventola, 1996). Las nominalizaciones estilísticas son, pues, marcadores complejos del discurso especializado. En (5)

> (5) Her distance from the sun, like that of Ceres, is about 263 millions of miles, and she **performs her revolution** about it, nearly in the same time (Phillips 1817: 68; énfasis añadido).

el autor podría haber recurrido a una estructura más simple por medio de una oración con verbo conjugado. Sin embargo, la elección de complejidad estructural por medio de una colocación delimita el nivel de especialización del autor, el público y el texto.

4. **Términos**: este tipo de nominalizaciones está más cerca de la codificación semántica de entidades como sustantivos. Los *términos* son dispositivos cognitivos que creamos y utilizamos para estudiar la realidad mediante el establecimiento de un conjunto de diferencias y fronteras (Calvin, 1996; von Eckardt, 1993; Lakoff, 1980). Son especialmente útiles en las disciplinas científicas, ya que proporcionan rastros semánticos de entidad (reificación) en procesos y entidades (Banks, 2005b). Resultan por lo tanto directrices funcionales que facilitan la organización de la información en la mente del lector. En (6)

> (6) But independent of theſe conſiderations, this rude ſyſtem was ſoon found incapable of ſtanding the teſt of **obſervation** and **experiment** (Bonnycastle, 1786, p. 59; énfasis añadido).

tanto *observation* como *experiment* se presentan como procesos plenamente reificados. La atención se centra en el proceso en sí, y omite toda la información acerca de agentes fácilmente reconocibles dado nuestro conocimiento compartido del mundo, así como objetos y circunstancias que se desconocen. Al poner de relieve a través de estas nominalizaciones dos puntos de enfoque claramente reconocibles, el autor ha proporcionado pautas funcionales sobre la manera de descodificar el texto.

El objetivo de esta tesis doctoral es la realización de un análisis de las nominalizaciones deverbales formadas por sufijación en un corpus de textos científicos escritos en inglés en los siglos XVIII y XIX con el fin de determinar en qué medida

éstas pueden ser consideradas marcadores del discurso científico. Esta línea de investigación es parte del estudio de la evolución histórica de inglés para fines específicos. Este objetivo principal se divide en cinco objetivos que se describen a continuación :

1. Estudio de nominalizaciones como marcadores del discurso científico: las nominalizaciones son estructuras complejas que por lo general contribuyen a aumentar el grado de ambigüedad en los textos. Este estudio analiza en profundidad no sólo la morfosintaxis, sino también las implicaciones funcionales de las nominalizaciones en los textos.
2. Análisis de los documentos de las nominalizaciones según criterios diacrónicos: siguiendo esta directriz pretendo establecer la evolución de las nominalizaciones durante los doscientos años que abarca el estudio .
3. Análisis según variables extralingüísticas: incluye la cronología, el sexo del autor, el lugar de educación y el tipo de texto. Esta información puede ayudar a esclarecer la relación entre el lenguaje y la sociedad en su dimensión histórica. Del mismo modo su aplicación a los resultados de los datos proporcionará información sobre cómo los factores sociológicos pueden ser causantes del cambio lingüístico.
4. Análisis en función de variables lingüísticas: el estudio de la estructura de las nominalizaciones se aborda a partir de un análisis etimológico de raíces y sufijos. Debido a que las nominalizaciones siempre actúan como núcleo de la frase nominal en la que aparecen, todos los elementos de la frase se analizan con la expectativa subyacente de que éstos incluyan información sobre el proceso . Por último, el análisis sintáctico pretende arrojar alguna luz sobre las funciones dentro de los textos .

5. Análisis de las nominalizaciones según tipología: ayudará a aclarar que las nominalizaciones no son compartimentos cerrados. La tipología desarrollada para este estudio contempla que la expresión lingüística de los procesos puede adquirir un cierto grado de características semánticas y léxicas de sustantivos o verbos. La aplicación de las variables lingüísticas a cada una de las tipologías descritas pretende delimitar diferencias formales y funcionales de las mismas.
6. Análisis sociohistórico de la ciencia en la época moderna: aunque este no es uno de los objetivos principales, el estudio de la ciencia en la época moderna y en especial la situación de exclusión de las mujeres científicas constituyen un foco importante del análisis.

Esta tesis está dividida en cuatro capítulos principales enmarcados por una introducción y un capítulo de conclusiones. Los cuatro capítulos centrales profundizan sobre los aspectos lingüísticos y extralingüísticos más relevantes en la historia de la lengua inglesa en el período moderno tardío y ponen especial énfasis en el uso de las nominalizaciones en el registro científico. El primer capítulo se inicia con un recuento de las causas y consecuencias de la revolución científica y su efecto sobre el cambio lingüístico y el desarrollo del registro científico en inglés en los siglos XVIII y XIX. Se proporciona un breve análisis del método científico ya que las prácticas científicas y el lenguaje científico utilizado en el siglo XVIII se derivan directamente del marco teórico establecido el siglo anterior. También se analiza el proceso de institucionalización de la ciencia porque se considera un factor de suma importancia para la creación y estandarización del registro científico. Este capítulo también contiene un análisis del papel de las mujeres científicas en estos siglos, ya que la variable del género de autor ha

supuesto una de las variables de estudio extralingüísticas. Para lograr una pequeña contextualización el primer capítulo también incluye una breve reseña de la situación de la disciplina de la astronomía en la época. La última sección de este capítulo aborda la situación del inglés en el período moderno, así como el surgimiento y la evolución del inglés como lengua de la ciencia.

El segundo capítulo está dedicado a la descripción de las consideraciones teóricas sobre el estudio de la nominalización. Aunque una gran parte de este capítulo consiste en una revisión bibliográfica de todas las teorías sobre la nominalización proporcionada por distintas escuelas lingüísticas, he puesto un especial énfasis en crear un enfoque original. El capítulo comienza con una reflexión sobre la definición de la nominalización y los problemas asociados. Estrechamente relacionado con los problemas de definición se encuentra el concepto de transferencia, lo que para muchas escuelas lingüísticas es uno de los rasgos definitorios de la nominalización. Se aborda el estudio desde dos ángulos diferentes: por un lado, la descripción de las principales teorías sobre la morfosintaxis y por otro, el estudio de las implicaciones funcionales. Después de analizar forma y función, se proporciona un comentario especial sobre el papel de las nominalizaciones como marcadores del discurso, antes de pasar a la descripción de la tipología creada para este estudio.

El tercer capítulo presenta una descripción del corpus de textos utilizados para el análisis, así como la metodología utilizada. La descripción del corpus incluye temas como el tamaño y la categorización textual, así como el sexo, la ocupación y la procedencia de los autores. De igual manera, también he incluido una breve descripción de la *CCT*, el motor de búsqueda utilizado para recuperar información del corpus. En la ultima parte del capítulo se presenta la metodología utilizada para este estudio: se

explican las búsquedas realizadas en el corpus, el proceso de desambiguación y la creación de la base de datos utilizada para el análisis. De igual manera, se detallan las variables de estudio utilizadas que incluyen tanto variables extralingüísticas (cronología, sexo del autor, lugar de educación del autor y tipo de texto) como intralingüísticas (estructura de las nominalizaciones y de sus frases, uso de sufijos, etimología, modificadores, construcciones posesivas, eliminación del agente, transformación de valencias verbales y complementos circunstanciales y función sintáctica)

El cuarto capítulo constituye el análisis de los datos obtenidos después de la explotación del corpus. Las variables que se describen en el tercer capítulo se aplican primero al número total de nominalizaciones que se encuentran en el corpus (8.446) y luego a cada uno de las cuatro tipologías descritas en el segundo capítulo. El análisis se ha llevado a cabo desde dos ángulos diferentes. En primer lugar todas las nominalizaciones se enfrentan a las variables extralingüísticas y lingüísticas. A continuación se presenta un análisis exclusivamente lingüístico de las nominalizaciones según las cuatro tipologías descritas en el capítulo dos. La aplicación de las variables extralingüísticas al estudio tipológico no mostró resultados significativos por lo que he preferido excluirlo de este capítulo, para no complicar la presentación de los resultados. La aplicación de las variables lingüísticas, a toda la serie de nominalizaciones y a cada una de las tipologías, se centra en la distinción entre forma y función tanto de las nominalizaziones en su forma léxica así como de las frases nominales en las que éstas funcionan como núcleos. Por lo tanto, el análisis de la nominalización como una unidad léxica trata en primer lugar aspectos sobre su morfología y luego aborda cuestiones relacionadas con su función. El análisis de las frases nominales también emula la

distinción entre forma y función, ya que incluye no sólo un estudio de los elementos sino también de las funciones y la semántica de esos elementos dentro de la frase nominal dirigida por la nominalización.

Por último, en el capítulo cinco se ofrecen las conclusiones generales y las futuras líneas de trabajo. Mi objetivo es que la base proporcionada por los datos cuantitativos obtenidos del análisis de corpus ayude a explicar el desarrollo de las nominalizaciones como marcadores del discurso científico en inglés en los siglos XVIII y XIX. Este es sin duda un tema muy interesante que puede tener un impacto directo en nuestra comprensión tanto de las nominalizaciones como en la forma de comunicar la ciencia. Espero que las conclusiones a las que he llegado después del análisis de datos pueden inspirar futuros trabajos sobre el papel de las nominalizaciones en el registro científico.

Fuentes primarias

Bonnycastle, John (1786). *An introduction to astronomy in a series of letters*. London: J. Johnson.

Olmstead, Denison (1841). *Letters on astronomy, addressed to a lady*. Boston: Marsh, Capen, Lyon, and Webb.

Philips, William (1817). *Eight familiar lectures on astronomy. An introduction to science for the use of young persons and others not conversant with the mathematics.* New York: J. Eastburn & Co.

Vince, Samuel (1790). *A treatise on practical astronomy*. Cambridge: J. Merrill, J. Nicholson & W. Lunn.

Watts, Isaac (1726). *The knowledge of the heavens and the earth made easy: Or, the first principles of astronomy and geography explain'd by the use of globes and maps*: *With a solution of the common problems by a plain* scale and Compasses as well as by the globe. London: J. Clark, R. Hett & R. Ford.

Whiston, William (1715). *Astronomical lectures* (I). London: R. Senex & W. Taylor.

Referencias citadas

Banks, David (2001). Analyse des discours spécialisés: Le domaine anglais. *Revue Française de Linguistique Appliquée, VI*(2), 7-16.

Banks, David (2005a). Emerging scientific discourse in the late seventeenth century. *Functions of Language, 12*(1), 65-86.

Banks, David (2005b). On the historical origins of nominalized process in scientific text. *English for Specific Purposes, 24*(3), 347-357.

Billig, Michael (2008). The language of critical discourse analysis: The case of nominalization. *Discourse Society, 19*, 783-800.

Blakemore, Diane (1987). *Semantic constraints on relevance.* Oxford: Blackwell.

Calvin, William H. (1996). *Como piensan los cerebros*. Madrid: Editorial Debate.

Crespo, Begoña (2004). General survey of the growth of scientific culture. A historical approach. In Elizabeth Woodward et al. (Eds.), *About culture* (pp. 57-65). A Coruña: Universidade da Coruña.

Crespo, Begoña (2011). Persuasion markers and ideology in eighteenth century philosophy texts (CEPhiT). *Revista de Lenguas para Fines Específicos, 14/15*, 199-228.

van Dijk, Teun A. (2006). Ideology and discourse analysis. *Journal of Political Ideologies, 11*(2), 115-140.

van Dijk, Teun A. (2008). Critical discourse analysis and nominalization: Problem or pseudo-problem? *Discourse Society, 19*, 821-828.

Downing, Angela (1997). Encapsulating Discourse Topics. *Estudios Ingleses de la Universidad Complutense, 5*, 147-168.

von Eckardt, Barbara (1993). *What is cognitive science?* Cambridge, MA: The MIT Press.

Fairclough, Norman (1992). *Discourse and social change*. Cambridge: Blackwell publishers.

Fowler, Roger (1991). *Language in the news: discourse and ideology in the press.* London & New York: Routledge.

Guillén Calve, Ignacio (1998). The textual interplay of grammatical metaphor on the nominalizations occurring in written medical English. *Journal of Pragmatics, 30*, 363-385.

Halliday, Michael A. K. (1985). *An introduction to functional grammar.* London: Continuum.

Halliday, Michael A. K. (2004). On the language of physical science. In Jonathan Webster (Ed.), *The language of physical science* (pp. 162-178). London: Continuum.

Lakoff, George & Johnson, Mark (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Moskowich Isabel, Lareo Inés, Camiña Rioboo Gonzalo & Crespo, Begoña. (Compilers). (2012). *A corpus of English texts on astronomy (CETA)* (CD ROM). Amsterdam/Philadelphia: John Benjamins.

Ventola, Eija (1996). Packing and unpacking of information in academic texts. In Eija Ventola & Anna Mauranen (Eds.), *Academic writing, intercultural and textual issues* (pp. 153-194). Amsterdam: John Benjamins.

Wodak, Ruth & Meyer, Michael (Eds.). (2001). *Methods of critical discourse analysis*. London: Sage Publications.

Appendix B

Doutoramento Internacional (2014)

Programa: Novas aproximacións aos estudos ingleses

Departamento de Filoloxía Inglesa

Universidade da Coruña

Directora: Begoña Crespo García

RESUMEN DA TESE DE DOUTORAMENTO

ON HOW THE MOTION OF THE STARS CHANGED THE LANGUAGE OF SCIENCE: A CORPUS-BASED STUDY OF DEVERBAL NOMINALIZATIONS IN ASTRONOMY TEXTS FROM 1700 TO 1900.

DE COMO *O MOVEMENTO DAS ESTRELAS* CAMBIOU A LINGUAXE DA CIENCIA: UN ESTUDO BASADO EN CORPUS DAS NOMINALIZACIÓNS DEVERBAIS EN TEXTOS DE ASTRONOMÍA ESCRITOS ENTRE 1700 E 1900.

Iria María Bello Viruega

A ciencia e a linguaxe da ciencia son dúas entidades inseparables. A dificultade da linguaxe científica non está limitada ao nivel léxico, senon que tamén se aplica a un número de estruturas gramaticais específicas (Halliday, 2004). As nominalizacións, un dos máis coñecidos marcadores do discurso científico, son estruturas complexas que codifican procesos en substantivos. Ao nominalizar, a cotío unha gran cantidade de información semántica se elimina, o que aumenta o grao de ambigüidade e a dificultade para descodificar correctamente a frase.

O obxectivo deste estudo é abordar o uso de nominalizacións formadas por sufixación en textos de rexistro científico escritos en inglés nos séculos XVIII e XIX. Neste período, varios cambios sociais afectaron a configuración da ciencia en Europa e isto tivo un impacto sobre a linguaxe. Os empiristas estableceron unha serie de orientacións sobre morfosintaxe, vocabulario especializado e estrutura do texto para a presentación de documentos observacionais ou experimentais. Estas pautas estilíticas convertíronse nun estándar de escritura tras a fundación da Royal Society (Crespo, 2004, 2011). As nominalizacións convertíronse nuns destes marcadores distintivos e, en consecuencia, a súa complexidade e a súa frecuencia aumentaron constantemente desde entón.

O material de corpus para este estudo foi tirado *do Corpus of English Texts on Astronomy (CETA)* (Moskowich et al., 2012). Este é un dos subcorpus do Coruña Corpus, *A Collection of Samples for the Historical Study of English Scientific Writing*. *CETA* contén dous textos por década e cada mostra contén preto de 10.000 palabras, o que fai un total de 800.000 palabras analizables. Outra ferramenta utilizada foi a *Coruña Corpus Tool* (*CCT*), unha ferramenta para recuperación de información especialmente deseñada para o seu uso en conxunto co corpus.

Entendo a nominalización como unha expresión lingüística dunha representación conceptual dun proceso ou estado de cousas nunha forma nominal. Esta definición baséase nas premisas establecidas por Downing (1997 p. 147), que considera que as situacións e os procesos poden ser expresados lingüisticamente de dous xeitos principais: cláusulas e nominalizacións. As linguas naturais teñen moitos recursos para expresar ideas similares. Esta flexibilidade, que pode ter matices importantes de significado, segue atraendo a atención de lingüistas e falantes, que ven reflexada a

riqueza da linguaxe. Como resultado, os procesos poden ser expresados por nominalizacións, como en (1)

> (1) From whence it is gathered, that **the apparent progreſſive Motion of the Fixed Stars** hath gone forward one Degree towards the conſequent Signs, in about Seventy Years ſpace (Whiston, 1715, p. 14; énfase engadido).

ou por medio de cláusulas con verbos conxugados , como en (2)

> (2) Aſtronomers know that not only the 12 Conſtellations of the Zodiac, but alſo **all the fix'd Stars move from the Weſt toward the Eaſt** about 50" in a Year, or one Degree in 71 Years, in Circles parallel to the Ecliptick (Watts, 1726, p. 34; énfase engadido).

Obviamente, aínda que ambas opcións transmiten o mesmo significado en liñas xenerais, cada unha destas dúas codificacións lingüísticas ten unha estrutura diferente e cumpre funcións distintas no texto. En (2) o verbo *move* controla a sintaxe da frase enteira a través dun sistema de valencias obrigatorias e modificadores opcionais. Así, o axente (*Conſtellations of the Zodiac, but alſo all the fix'd Stars*) aparece como suxeito. A dirección do movemento (*from the Weſt toward the Eaſt*) faise explícita e tamén aparece información sobre cómo (i*n Circles parallel to the Ecliptick*) y cánto (*about 50" in a Year, or one Degree in 71 Years*) as constelacións e as estrelas se moven. Do mesmo xeito, en (1) a nominalización *motion* exerce control sobre a súa frase, a cal insértase a súa vez nunha oración máis grande. A estrutura non é tan ríxida neste caso, xa que, por definición, todos os elementos da frase nominal agás o núcleo son opcionais. Isto permite unha disposición máis complexa dos elementos. Así, en (1) a información sobre o axente (*of the Fixed Stars*), e cómo (*progessive*) o movemento se leva a cabo encóntrase dentro dos modificadores nomináis. Isto, porén, non implica a

reducción da cantidade de información xa que na frase verbal tamén hai información sobre a dirección (*towards the conſequent Signs, in about Seventy Years ſpace*) e a duración (*one Degree)* do movemento.

Funcionalmente, a expresión do proceso como un verbo en (2) sitúase máis cerca da experienza do falante en termos de secuenciación e experiencia da realidade xa que toda a información sobre proceso mantense cerca do verbo. Con todo, a configuración a través dunha nominalización permite presentar o proceso non como un simple relato da realidade, senon como consecuencia do discurso anterior que pasa a ser entendido como un obxecto, dada a súa codificación como un substantivo.

A funcionalidade das nominalizacións foi o punto central de investigación de moitos autores, sobre todo dentro da escola sistémico funcional (Bancos, 2005a e 2005b, Guillén, 1998; Halliday, 1985, 2004; Ventola, 1996). Unha breve recensión das principais funcións desempeñadas pola nominalización en textos pode incluír: cohesión léxica (repetición e resumo); economía, concisión, a condensación da información e contextualización da información (en relación coas estruturas de información) que se traduce no avance do discurso. O dinamismo que as nominalizacións confiren a estrutura temática dos textos adoita contrapoñerse aos efectos da eliminación do axente, un tema amplamente discutido pola escola de análise crítica do discurso (Billig, 2008; van Dijk, 2006, 2008; Fairclough, 1992; Fowler, 1991; Wodak e Meyer, 2001). As nominalizacións poden silenciar información sobre os axentes e as circunstancias, o que pode levar a certos efectos ideolóxicos. Con todo, dende un punto de vista cognitivo, as nominalizacións responden a un modelo mental que o escritor fixo sobre o proceso que quere comunicar e o contexto do proceso de comunicación. Neste sentido, convírtense en pautas funcionáis para o procesamento da información xa que proporcionan

instrucións e limitan a posible ambigüidade das frases, aforrando esforzos no procesamento (Blakemore, 1987). Para este estudo, desenvolvín unha tipoloxía das nominalizacións que ten en conta características formais e funcionais:

1. **Nominalización condicionada:** polo xeral, ven dictada polas necesidades gramaticais do texto. Extremadamente concisa, é xeralmente a opción que os escritores utilizan para condensar varios procesos nunha sola frase e omitir elementos innecesarios. En (3)

> (3) The same comet, also, came very near the earth; so that, had its quantity of matter been equal to that of the earth, it would, by its attraction, have caused the earth to revolve in an orbit so much larger than at present, (...) (Olmstead 1841: 318; énfase engadido).

o autor quere presentar o proceso (atracción) como axente causante de outro proceso (que a Terra xira nunha órbita). Debido a que a información sobre o axente, o obxecto e as circunstancias en que o proceso de atracción ten lugar son facilmente extraíbles do contexto e derívanse do noso coñecemento compartido do mundo, o escritor prefiriu centrar a atención do lector no proceso en sí mesmo e presentalo a través dunha nominalización.

2. **Nominalización temática:** cumpre con todas as características descritas polos funcionalistas (Banks 2001, 2005a , 2005b): un proceso codificado nun grupo verbal (codificación congruente) e que funciona como rema pódese convertir nunha nominalización (metáfora gramatical) no tema da seguinte frase. Tamén serve como un dispositivo de cohesión, repite e resume a información e ofrece un punto de partida que facilita o avance do discurso. En (4)

> (4) but if the limbs be ſeperated, the object end is inclined to the quadrant, and muſt be **adjuſted** accordingly, and repeat the operation till the limbs coincide at both wires, and the **adjuſtment** is made (Vince, 1790, p. 15; énfase engadido).

o avance do discurso consíguese convertindo o verbo *adjusted* na nominalización *adjustment* ao final da oración. Semanticamente, este tipo de nominalización presenta una gran similitude coa codificación dos procesos a través de cláusulas con verbo conxugado.

3. **Nominalización estilística**: a elección da codificación nominal responde, nesta tipoloxía, a criterios estilísticos. A complexidade estilística é considerada unha forma de "codificación gremial", un código que só os membros dunha comunidade dominan e que diferencia aos membros experimentados dos novatos e do resto da sociedade (Ventola, 1996). As nominalizacións estilísticas son, polo tanto, marcadores complexos do discurso especializado. En (5)

> (5) Her distance from the sun, like that of Ceres, is about 263 millions of miles, and she **performs her revolution** about it, nearly in the same time (Phillips 1817: 68; énfase engadido).

o autor podería recurrir a unha estrutura máis simple a través dunha oración co verbo conxugado. Con todo, a elección de complexidade estrutural a través dunha colocación delimita o nivel de coñecemento do autor, o público e o texto.

3. **Térmos:** este tipo de nominalizacións está máis preto da codificación semántica das entidades como substantivos. Os *termos* son dispositivos cognitivos que creamos e utilizamos para estudar a realidade a través do establecimento dun conxunto de diferenzas e fronteras (Calvin, 1996; von Eckardt, 1993; Lakoff, 1990). Son

especialmente útiles nas disciplinas científicas, xa que proporcionan rastros semánticos de entidade (reificación) nos procesos e entidades (Banks, 2005b). Resultan polo tanto directrices funcionais que facilitan a organización da información na mente do lector. En (6)

> (6) But independent of theſe conſiderations, this rude ſyſtem was ſoon found incapable of ſtanding the teſt of **obſervation** and **experiment** (Bonnycastle, 1786, p. 59; énfase engadido).

tanto *observation* como *experiment* preséntanse como procesos totalmente reificados. A atención céntrase no proceso en sí, e omite toda a información sobre axentes facilmente recoñecibles dado o noso coñecemento compartido do mundo, así como obxectos e circunstancias que se descoñecen. Ao destacar puntos de enfoque a través destas dúas nominalizacións, o autor proporciona pautas funcionais sobre a forma de decodificar o texto.

O obxectivo desta tese doutoral é a realización dunha análise das nominalizacións deverbais formadas por sufixación nun corpus de textos científicos escritos en inglés, nos séculos XVIII e XIX, a fin de determinar en qué medida estas poden ser consideradas marcadores do discurso científico. Esta liña de investigación forma parte do estudo da evolución histórica do inglés para fins específicos. O obxectivo principal é dividido en cinco obxectivos que se describen a continuación:

1. Estudo de nominalizacións como marcadores do discurso científico: as nominalizacións son estruturas complexas que normalmente contribúen a aumentar o grao de ambigüidade nos textos. Este estudo examina en profundidade non só a morfosintaxe, mais tamén as implicacións funcionais de nominalizacións nos textos.

2. Análise das nominalizacións segundo criterios diacrónicos: seguindo esta directriz pretendo establecer a evolución das nominalizacións, proceso durante os douscentos anos cubertos polo estudo.
3. Análise segundo variables extralinguísticas: inclúe a cronoloxía, o sexo do autor, o lugar de educación e o tipo de texto. Esta información pode axudar a aclarar a relación entre a linguaxe e a sociedade na súa dimensión histórica. Do mesmo modo a súa aplicación aos resultados dos datos proporcionará información sobre cómo os factores sociolóxicos poden ser causantes do cambio lingüístico.
4. Análise en base a variables lingüísticas: o estudo da estrutura das nominalizacións abórdase dende unha análise etimolóxica de raíces e sufixos. Debido a que as nominalizaións sempre actúan como o núcleo da frase nominal na que aparecen, todos os elementos da frase analízanse coa fin subxacente de que estes inclúen información sobre o proceso. Finalmente, a análise sintáctica ten como obxectivo dar a coñecer as funcións das nominalizacións dentro dos textos.
5. Análise tipolóxica das nominalizacións: axudará a aclarar que as nominalizacións non son compartimentos pechados. A tipoloxía plsntexada para este estudo considera que a expresión lingüística dos procesos pode adquirir un certo grao de trazos semánticos e léxicos de substantivos ou verbos. A aplicación das variables lingüísticas para cada unha das tipoloxías descritas pretende delimitar diferenzas formais e funcionais das mesmas.
6. Análise sociohistórica da ciencia na era moderna: aínda que este non é un dos obxectivos principais, o estudo da ciencia na era moderna e, en especial a situación de exclusión das mulleres científicas constitúe un foco importante de análise.

Esta tese está dividida en catro capítulos principais enmarcados por unha introdución e un capítulo de conclusións. Os catro capítulos principais afondan sobre os aspectos lingüísticos e extralingüísticos máis relevantes na historia da lingua inglesa no período moderno tardío e poñen unha énfase especial no uso das nominalizacións no rexistro científico. O primeiro capítulo comeza cun relato das causas e consecuencias da revolución científica e o seu efecto no cambio lingüístico e o desenvolvemento do rexistro científico en inglés, nos séculos XVIII e XIX. Proporciónase unha breve análise do método científico xa que as prácticas científicas e a linguaxe científica utilizada no século XVIII derívanse directamente do marco teórico establecido no século anterior. Tamén se analiza o proceso de institucionalización da ciencia pois é un factor de suma importancia para a creación e estandarización do rexistro científico. Este capítulo tamén contén unha análise do papel das mulleres científicas nestes séculos, tendo en conta que a variable do xénero do autor supuxo unha das variables de estudo extralinguístico. Para obter unha pequena contextualización o primeiro capítulo tamén inclúe unha breve recensión da situación da astronomía nesta época. A última sección deste capítulo trata da situación do inglés no período moderno, e da aparición e evolución do inglés como lingua de ciencia.

O segundo capítulo está dedicado á descrición das consideracións teóricas sobre o estudo da nominalización. Aínda que unha gran parte deste capítulo consiste nunha revisión bibliográfica de todas as teorías sobre nominalización proporcionada por distintas escolas lingüísticas, puxen especial énfase en crear un enfoque orixinal. O capítulo comeza cunha reflexión sobre a definición da nominalización e os problemas asociados. Intimamente relacionado cos problemas de definición atópase o concepto de transferencia, o que para moitas escolas lingüísticas é unha das características que

definen a nominalización. Trátase o estudo dende dous ángulos diferentes: primeiro, a descrición das principais teorías sobre morfosintaxe e en segundo lugar, o estudo das implicacións funcionais. Tras ter analizado a forma e a función, proporciónase un comentario especial sobre o papel das nominalizacións como marcadores do discurso, antes de abordar a descrición da tipoloxía creada para este estudo.

O terceiro capítulo presenta unha descrición do corpus de textos utilizados para a análise como a metodoloxía utilizada. A descrición do corpus inclúe temas como o tamaño, a categorización do texto como o sexo, a profesión e a orixe dos autores. Do mesmo xeito, tamén incluín unha breve descrición da *CCT*, a ferramenta empregada para recuperar información do corpus. A metodoloxía utilizada neste estudo é presentada na última parte do capítulo: explícanse as procuras feitas no corpus, o proceso de desambiguación e a creación da base de datos utilizada para a análise. Do mesmo xeito, detállanse as variables de estudo utilizadas que inclúen variables extralinguísticas (cronoloxía, sexo do autor, lugar de educación do autor e tipo de texto do autor) e intralingüísticas (estrutura das nominalizacións e das súas frases, uso de sufixos, etimoloxía, modificadores, construcións posesivas, eliminación do axente, transformación de valencias verbais e complementos circunstanciais e función sintática)

O cuarto capítulo constitúe a análise dos datos obtidos tras a explotación do corpus. As variables que foron descritas no terceiro capítulo aplicáronse primeiro ao número total de nominalizacións que se atopan no corpus (8.446) e logo, a cada unha das catro tipoloxías descritas no segundo capítulo. A análise foi realizada dende dous ángulos diferentes. Nun primeiro lugar todas as nominalizacións se enfrentan as variables extralingüísticas e lingüísticas. A continuación preséntase unha análise exclusivamente lingüística das nominalizacións de acordo coas catro tipoloxías descritas

no capítulo dous. A aplicación das variables extralinguísticas para o estudo tipolóxico non mostrou resultados significativos polo que preferín eliminalo deste capítulo, para non complicar a presentación dos resultados. A aplicación das variables lingüísticas, a toda unha serie de nominalizacións e a cada unha das tipoloxías, céntrase na distinción entre forma e función, tanto das nominalizaziones na súa forma léxica así como das frases nominais nas que estas funcionan como núcleos. Polo tanto, a análise da nominalización como unha unidade léxica trata en primeiro lugar aspectos sobre a súa morfoloxía e logo, aborda cuestións relacionadas coa súa función. A análise das frases nominais tamén emula a distinción entre forma e función, xa que inclúe non só un estudo dos elementos senón tamén das funcións e a semántica destes elementos dentro da frase nominal dirixida pola nominalización.

Finalmente, no capítulo cinco ofrécense as conclusións xerais e as futuras liñas de traballo. O meu obxectivo é que a base proporcionada polos datos cuantitativos obtidos da análise de corpus axude a explicar o desenvolvemento das nominalizacións como marcadores do discurso científico en inglés nos séculos XVIII e XIX. Este é sen dúbida un tema moi interesante que pode ter un impacto directo tanto na nosa comprensión das nominalizacións como na forma de comunicar a ciencia. Espero que as conclusións as que cheguei despois da análise de datos poidan inspirar futuros traballos sobre o papel das nominalizacións no rexistro científico.

Fontes primarias

Bonnycastle, John (1786). An introduction to astronomy in a series of letters. London: J. Johnson.

Olmstead, Denison (1841). Letters on astronomy, addressed to a lady. Boston: Marsh, Capen, Lyon, and Webb.

Philips, William (1817). Eight familiar lectures on astronomy. An introduction to science for the use of young persons and others not conversant with the mathematics. New York: J. Eastburn & Co.

Vince, Samuel (1790). *A treatise on practical astronomy*. Cambridge: J. Merrill, J. Nicholson & W. Lunn.

Watts, Isaac (1726). The knowledge of the heavens and the earth made easy: Or, the first principles of astronomy and geography explain'd by the use of globes and maps: With a solution of the common problems by a plain scale and Compasses as well as by the globe. London: J. Clark, R. Hett & R. Ford.

Whiston, William (1715). *Astronomical lectures* (I). London: R. Senex & W. Taylor.

Referencias citadas

Banks, David (2001). Analyse des discours spécialisés: Le domaine anglais. *Revue Française de Linguistique Appliquée, VI*(2), 7-16.

Banks, David (2005a). Emerging scientific discourse in the late seventeenth century. *Functions of Language, 12*(1), 65-86.

Banks, David (2005b). On the historical origins of nominalized process in scientific text. *English for Specific Purposes, 24*(3), 347-357.

Billig, Michael (2008). The language of critical discourse analysis: The case of nominalization. *Discourse Society, 19*, 783-800.

Blakemore, Diane (1987). Semantic constraints on relevance. Oxford: Blackwell.

Calvin, William H. (1996). *Como piensan los cerebros*. Madrid: Editorial Debate.

Crespo, Begoña (2004). General survey of the growth of scientific culture. A historical approach. In Elizabeth Woodward et al. (Eds.), *About culture* (pp. 57-65). A Coruña: Universidade da Coruña.

Crespo, Begoña (2011). Persuasion markers and ideology in eighteenth century philosophy texts (CEPhiT). *Revista de Lenguas para Fines Específicos, 14/15*, 199-228.

van Dijk, Teun A. (2006). Ideology and discourse analysis. *Journal of Political Ideologies, 11*(2), 115-140.

van Dijk, Teun A. (2008). Critical discourse analysis and nominalization: Problem or pseudo-problem? *Discourse Society, 19*, 821-828.

Downing, Angela (1997). Encapsulating Discourse Topics. *Estudios Ingleses de la Universidad Complutense, 5*, 147-168.

von Eckardt, Barbara (1993). *What is cognitive science?* Cambridge, MA: The MIT Press.

Fairclough, Norman (1992). *Discourse and social change*. Cambridge: Blackwell publishers.

Fowler, Roger (1991). Language in the news: discourse and ideology in the press. London & New York: Routledge.

Guillén Calve, Ignacio (1998). The textual interplay of grammatical metaphor on the nominalizations occurring in written medical English. *Journal of Pragmatics, 30*, 363-385.

Halliday, Michael A. K. (1985). An introduction to functional grammar. London: Continuum.

Halliday, Michael A. K. (2004). On the language of physical science. In Jonathan Webster (Ed.), *The language of physical science* (pp. 162-178). London: Continuum.

Lakoff, George & Johnson, Mark (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Moskowich Isabel, Lareo Inés, Camiña Rioboo Gonzalo & Crespo, Begoña. (Compilers). (2012). *A corpus of English texts on astronomy (CETA)* (CD ROM). Amsterdam/Philadelphia: John Benjamins.

Ventola, Eija (1996). Packing and unpacking of information in academic texts. In Eija Ventola & Anna Mauranen (Eds.), *Academic writing, intercultural and textual issues* (pp. 153-194). Amsterdam: John Benjamins.

Wodak, Ruth & Meyer, Michael (Eds.). (2001). *Methods of critical discourse analysis*. London: Sage Publications.